

METRON

RIVISTA INTERNAZIONALE DI STATISTICA — REVUE INTERNATIONALE DE STATISTIQUE
INTERNATIONAL REVIEW OF STATISTICS — INTERNATIONALE STATISTISCHE ZEITSCHRIFT
REVISTA INTERNACIONAL DE ESTADISTICA

DIRETTORE — DIRECTEUR — EDITOR — HERAUSGEBER — DIRECTOR

Dott. Corrado Gini, *Prof. emerito della Università di Roma*

COMITATO DIRETTIVO — COMITE' DE DIRECTION — EDITORIAL COMMITTEE
DIREKTION-KOMITEE — COMITE' DIRECTIVO

E.P. Billeter, Fribourg (Suisse); **F.P. Cantelli**, Roma (Italia); **V. Castellano**, Roma (Italia);
C. Dagum, Cordoba (Argentina); **M.P. Geppert**, (Miss.), Frankfurt a.M. (Deutschland);
B. Gil, Jerusalem (Israel); **L. Savage**, Chicago (U.S.A.);
H. von Schelling, Schenectad, New York (U.S.A.); **P. Steriotis**, Athens (Greece);
P. Whittle, Wellington (New Zealand); **S. Wilks**, Princeton (U.S.A.)
H. Wold, Stockholm (Sweden); **R. Yüceulug**, Ankara (Turkey).

SEGRETARI DI REDAZIONE — SECRÉTAIRES DE RÉDACTION
EDITORIAL SECRETARIES — REDAKTIONSSSEKRETAERE — SECRETARIOS DE LA REDACCIÓN

Proff. C. Benedetti, T. Salvemini

Vol. XX - N. 1-4

31-XII-1960

SOMMARIO-SOMMAIRE-CONTENTS-INHALT-SUMARIO

O. M. J. Mittmann. <i>Eine Methode zur Auffindung ursächlicher Zusammenhänge mittels partieller Korrelationskoeffizienten</i>	pag. 3
Camilo Dagum. <i>Teoría de la transvariación. Sus aplicaciones a la Economía</i>	» 45
Vittorio Castellano. <i>Sull'insieme delle distribuzioni doppie e triple risultanti dall'associazione una a una delle unità di due o tre distribuzioni semplici aventi lo stesso numero di unità</i>	» 251
D. P. Banerjee. <i>On the forms of some invariants of probability distribution</i>	» 299
Carlo Benedetti. <i>Su alcune singolari questioni statistiche ed econometriche connesse con le medie</i>	» 307
Corrado Gini. <i>Statistical Methods with special reference to Agriculture</i>	» 319
Bibliografia	» 421

DIREZIONE E REDAZIONE - DIRECTION ET REDACTION - EDITORIAL OFFICE -
DIREKTION UND REDAKTION - DIRECCIÓN Y REDACCION
10, Via delle Terme di Diocleziano - Roma (ITALIA)

AMMINISTRAZIONE - ADMINISTRATEUR - MANAGER - VERWALTER - ADMINISTRACIÓN
Istituto di Statistica - Facoltà di Scienze Statistiche - Città Universitaria - Roma (ITALIA)

ARTICOLI GIUNTI ALLA RIVISTA CHE
VERRANNO PUBBLICATI NEI PROSSIMI
NUMERI (*secondo l'ordine di arrivo*).

ARTICLES RECEIVED BY THE REVIEW
WHICH WILL BE PUBLISHED IN FU-
TURE ISSUES (*according to date of
receipt*).

ARTIKEL DIE AN DIE ZEITSCHRIFT ANGELANGT SIND UND WELCHE IN
DEN FOLGENDEN NUMMERN ERSCHEINEN WERDEN (*nach der Reihenfolge des
Eingangs*).

ARTICLES REÇUS PAR LA REVUE
ET À PARAÎTRE PROCHAINEMENT
(*d'après la date de reception*).

ARTICULOS LLEGADOS A LA REVISTA
QUE SE PUBLICARÁN EN LOS PROXIMOS
NUMEROS (*según el orden de su llegada*).

C. GINI e A. TREZZA, *Anormalità delle distribuzioni statistiche dei
caratteri antropologici*.

ERNST P. BILLETER, *Automation und Wirtschaftswissenschaft*.

A. HERZEL, *Nuovi contributi alla programmazione lineare*.

A. HERZEL, *Le tabelle di cograduazione e contrograduazione e la
programmazione lineare*.

A. HERZEL, *Su una limitazione a priori dei massimi e minimi ri-
cercati nei problemi di programmazione lineare*.

A. HERZEL, *Sul « problema del commesso viaggiatore »*.

G. LETI, *Le distribuzioni degli estremi e del campo di variazione
dei campioni di una popolazione discreta e finita*.

G. LETI, *Il termine generico delle tabelle di cograduazione e di con-
trograduazione*.

Gli Autori degli articoli inviati
per la pubblicazione nella Rivista
rinunciano in favore della medesima
alla proprietà letteraria degli artico-
li stessi, qualora vengano pubbli-
cati.

Les Auteurs des articles envoyés
à la Revue, pour y être publiés, re-
noncent, en faveur de celle-ci, à la
propriété littéraire de leurs articles,
s'ils sont acceptés.

Die Verfasser der zur Veröffent-
lichung in der Zeitschrift zugesand-
ten Aufsätze, werden, falls selbige
veröffentlicht werden, auf ihre Ver-
fasserrechte zu Gunsten der Zeit-
schrift verzichten müssen.

Los Autores de los artículos en-
viados para su publicación en « Me-
tron » renuncian su propiedad a
favor de la Revista cuando los ar-
tículos sean publicados.

The Authors of papers sent for publication in the Review are suppo-
sed to give up their copyright in favour of the Review if the papers
are published.

O.M.J. MITTMANN

**Eine Methode zur Auffindung
ursächlicher Zusammenhänge mittels partieller
Korrelationskoeffizienten**

EINLEITUNG

Wenn man in der Physik den Einfluss der Wärme auf die Ausdehnung eines Metall-Stabes untersucht, so erhält man eine Korrelations-Tabelle mit einer linearen Regressionslinie, wobei die Abweichungen von der Regressionsgeraden rein zufällig streuende Beobachtungsfehler darstellen. Dass es sich um einen Kausal-Zusammenhang zwischen der Länge des Metall-Stabes und der Wärme handelt, wird im Laufe der Zeit dadurch verifiziert, dass die gefundene lineare Abhängigkeit zwischen der Metallstab-Länge und der Temperatur immer wieder unter den verschiedensten irdischen Umwelt-Bedingungen bestätigt wird, d.h. dass zwischen den sonstigen auf der Erde vorkommenden Beobachtungs-Objekten und der Metallstab-Länge die Korrelation Null besteht.

Untersucht der Physiker die Fall-Gesetze, so findet er für die Fall-Strecke s und die während des Falls verflossene Zeit t eine Korrelations-Tafel mit der Regressionslinie $s = \frac{1}{2} g t^2$, wobei g eine Konstante ist. Später stellt er fest, dass die zurückgelegte Weg-Strecke zwar im wesentlichen von der Zeit t abhängt, aber der Luft-Widerstand auch noch eine Rolle spielt. — Ähnlich liegen die Dinge, wenn es sich um eine Geschoss-Bahn in der Ballistik handelt.

Genau genommen ist natürlich kein korrelativer Zusammenhang zwischen zwei Beobachtungs-Objekten ohne weiteres mit einem Kausal-Zusammenhang zu identifizieren. Denn es ist theoretisch auch im Falle sehr starker Korrelation immer möglich,

dass die beiden Beobachtungs-Objekte nur über eine dritte, vielleicht gänzlich unbekannte Grösse miteinander korreliert sind, d.h. dass kein Kausal-Zusammenhang, sondern ein indirekter korrelativer Zusammenhang vorliegt. Besteht eine aussergewöhnlich hohe Korrelation, wird man von vornherein geneigt sein, einen Kausal-Zusammenhang zu vermuten. Ist die Korrelation nicht besonders hoch, wird man skeptischer sein und sich vor Augen halten, dass ein korrelativer Zusammenhang ja nicht ohne weiteres ein Kausal-Zusammenhang zu sein braucht.

Die Frage, die sich nun erhebt, lautet: *Kann man auf statistischem Wege klären, ob ein korrelativer Zusammenhang nur ein indirekter ist oder ob ein direkter, kausaler Zusammenhang vorliegt?* Im Folgenden soll ein Weg aufgezeigt werden, der dem vom Physiker begangenen Weg ähnlich ist. Anstelle einer jahrelangen Verifizierung unter verschiedensten Umwelt-Bedingungen soll dabei freilich eine sofortige Ausweitung der Korrelations-Untersuchungen auf eine möglichst grosse Anzahl von Umwelt-Objekten erfolgen.

TEIL I

A. Monokausal-Theorie

I. Eine Ursache und 2 abhängige Variable. Wenn auch der Fall, dass ein Beobachtungs-Objekt nur mit einer einzigen Ursache zusammenhängt, einen idealen Sonderfall darstellt, so ist er doch theoretisch und praktisch bei Sachlagen wichtig, in denen eine Haupt-Ursache und daneben einige schwach wirkende Neben-Ursachen bedeutungsvoll sind. Seine Wichtigkeit erhöht sich, je weniger die Neben-Ursachen gegenüber der Haupt-Ursache ins Gewicht fallen.

Es seien 3 Zufalls-Variable a , b , u gegeben, zwischen denen die Korrelationen r_{ab} , r_{au} , r_{bu} bestehen; r sei der sogenannte Korrelationskoeffizient. Dann ist der partielle Korrelationskoeffizient zwischen a und b unter Ausschaltung des Einflusses von u gegeben durch

$$r_{ab,u} = \frac{r_{ab} - r_{au} r_{bu}}{\sqrt{(1 - r_{au}^2)(1 - r_{bu}^2)}}, \quad (I)$$

lineare Regression vorausgesetzt. Praktisch kann das Vorliegen linearer Regression immer durch Logarithmierung oder eine andere geeignete Koordinaten-Transformation erreicht werden.

Sofern zwischen a und b nur ein indirekter Zusammenhang über u besteht, d.h. $r_{ab,u} = 0$ ist, muss der Zähler in (1) verschwinden; es gilt dann also

$$r_{ab} = r_{au} r_{bu}. \quad (2)$$

Da r stets zwischen Null und $+1$ oder zwischen Null und -1 liegt ($|r| < 1$), ist Gleichung (2) nur möglich, wenn r_{ab} näher an Null liegt als r_{au} und als r_{bu} . Demnach gilt die Doppelungleichung

$$\begin{cases} |r_{ab}| < |r_{au}| \\ |r_{ab}| < |r_{bu}| \end{cases} \quad (3)$$

In Worten: Der indirekte Zusammenhang zwischen a und b ist schwächer als der direkte Zusammenhang zwischen a und u sowie der direkte Zusammenhang zwischen b und u .

Bilden wir die drei partiellen Korrelationskoeffizienten $r_{ab,u}$, $r_{au,b}$ und $r_{bu,a}$, dann verschwindet selbstverständlich bloss $r_{ab,u}$, wogegen $r_{au,b}$ und $r_{bu,a}$ von Null verschieden sind.

II. Eine Ursache und 3 abhängige Variable. Nicht selbstverständlich liegen die Dinge, sofern wir 3 Variable a, b, c haben, zwischen denen nur indirekte Zusammenhänge über eine Ursache u bestehen, und sofern wir die partiellen Korrelationskoeffizienten zwischen je 2 dieser 3 Variablen unter Ausschaltung der dritten Variablen berechnen. Bei alleinigem Bestehen indirekter Zusammenhänge zwischen a, b und c über eine Ursache u gilt

$$\begin{cases} r_{ab} - r_{au} r_{bu} = 0 \\ r_{ac} - r_{au} r_{cu} = 0 \\ r_{bc} - r_{au} r_{cu} = 0 \end{cases} \quad (4)$$

Der partielle Korrelationskoeffizient zwischen a und b unter Ausschaltung von c wird hiernach

$$r_{ab,c} = \frac{r_{ab} - r_{ac} r_{bc}}{\sqrt{(1 - r_{ac}^2)(1 - r_{bc}^2)}} = r_{ab} \frac{1 - r_{cu}^2}{\sqrt{(1 - r_{au}^2 r_{cu}^2)(1 - r_{bu}^2 r_{cu}^2)}},$$

was nur dann gleich Null sein kann, wenn entweder $r_{ab} = 0$ oder $r_{cu} = 1$. Da nach Voraussetzung weder $r_{ab} = 0$ noch $r_{cu} = 1$ ist, gilt demzufolge, dass

$$r_{ab,c} \neq 0. \quad (6)$$

In Worten: *Der partielle Korrelationskoeffizient zwischen Variablen, zwischen denen nur indirekte Zusammenhänge bestehen, ist stets von Null verschieden.* — Betrachten wir u nicht als einzelne Ursache, sondern als Ursachen-Komplex, so gilt der ausgesprochene Satz ganz allgemein.

B. Polykausal-Theorie

I. Frage der Identität zwischen maximal starkem und direktem Zusammenhang. Vorgegeben seien 2 Variable a und b , die nur indirekt über eine Ursache u und eine Ursache (oder einen Ursachen-Komplex) v zusammenhängen. Für alle auftretenden gewöhnlichen Korrelationskoeffizienten und partiellen Korrelationskoeffizienten erster Ordnung sei lineare Regression vorausgesetzt. Dann ist der partielle Korrelationskoeffizient zwischen a und b unter Ausschaltung von u und v in zweifacher Form schreibbar, nämlich

$$r_{ab,uv} = \frac{r_{ab,u} - r_{av,u} \cdot r_{bv,u}}{\sqrt{(1 - r_{av,u}^2)(1 - r_{bv,u}^2)}} = \frac{r_{ab,v} - r_{au,v} \cdot r_{bu,v}}{\sqrt{(1 - r_{au,v}^2)(1 - r_{bu,v}^2)}} = 0. \quad (7)$$

Wegen des Verschwindens von $r_{ab,uv}$ gilt

$$\begin{cases} r_{ab,u} = r_{av,u} \cdot r_{bv,u} \\ r_{ab,v} = r_{au,v} \cdot r_{bu,v} \end{cases} \quad (8)$$

Der gewöhnliche Korrelationskoeffizient zwischen a und b ergibt sich hieraus zu

$$r_{ab} = r_{au} \cdot r_{bu} + \frac{(r_{av} - r_{au} \cdot r_{uv})(r_{bv} - r_{bu} \cdot r_{uv})}{1 - r_{uv}^2}. \quad (9)$$

Während im Rahmen der Monokausal-Theorie die einfache Formel (2) besagt, dass $r_{ab} = r_{au} \cdot r_{bu}$ ist und infolgedessen $|r_{ab}| < |r_{au}|$

und $|r_{ab}| < |r_{bu}|$ sein muss, kann aus Formel (9) nicht geschlossen werden, dass $|r_{ab}|$ stets kleiner als die Korrelationskoeffizienten ist, welche direkte Zusammenhänge darstellen. Ein Beispiel möge das näher veranschaulichen. Wenn speziell $r_{uv} = 0$ ist, geht Gleichung (9) über in

$$r_{ab} = r_{au} \cdot r_{bu} + r_{av} \cdot r_{bv}, \quad (10)$$

und wenn dann z.B. $r_{au} = 0,70$ und $r_{bu} = r_{av} = r_{bv} = 0,60$ ist, wird $r_{ab} = 0,42 + 0,36 + 0,78$, d.h. r_{ab} wird hier grösser als jeder der Korrelationskoeffizienten, die einen direkten Zusammenhang ausdrücken.

Unter Bedingungen einer gewissen Gleichmässigkeit des Feldes der Korrelationskoeffizienten, die einen direkten Zusammenhang ausdrücken, lässt sich jedoch zeigen, dass der maximal starke Zusammenhang stets ein direkter Zusammenhang ist.

II. Hinreichende Bedingungen für Identität zwischen maximal starkem und direktem Zusammenhang. Dafür, dass der maximal starke Zusammenhang ein direkter Zusammenhang ist, lassen sich eine Unmenge von Fällen anführen. Es sind dies zwar lauter Spezialfälle, aber diese Spezialfälle stellen Sachlagen dar, die so ähnlich wie in der lebendigen Wirklichkeit gelagert sind.

Fall 1. Falls $r_{bu} = r_{av} = r_{uv} = r_{bv} = \vartheta r_{au}$ ist, geht (9) über in

$$r_{ab} = \vartheta r_{au}^2 \frac{1 + \vartheta}{1 + \vartheta r_{au}}, \quad (11)$$

und es wird $|r_{ab}| < |r_{au}|$, wenn

$$\vartheta^2 < \frac{1}{|r_{au}|}. \quad (12)$$

Sofern nun r_{au} der absolut grösste der Korrelationskoeffizienten ist, die einen direkten Zusammenhang ausdrücken, und ϑ zwischen 0 und +1 liegt, ist Ungleichung (12) stets erfüllt. Es ist demnach im Falle obiger Bedingungen stets $|r_{ab}| < |r_{au}|$.

Fall 2. Falls $r_{bu} = \vartheta^2 \cdot r_{au}$ und $r_{av} = r_{uv} = r_{bv} = \vartheta r_{au}$, dann geht (9) über in

$$r_{ab} = \vartheta^2 r_{au}^2 \frac{2 - \vartheta^2 \cdot r_{au} - r_{au}}{1 - \vartheta^2 \cdot r_{au}^2}, \quad (13)$$

und es wird $|r_{ab}| < |r_{au}|$, wenn

$$(\vartheta^2 \cdot r_{au} - 1)^2 > 0. \quad (14)$$

Sofern nun r_{au} wieder der absolut grösste der Korrelationskoeffizienten ist, die einen direkten Zusammenhang ausdrücken, und ϑ zwischen 0 und +1 liegt, ist Ungleichung (14) unter den Bedingungen des Falles 2 stets erfüllt. Es ist also dann stets $|r_{ab}| < |r_{au}|$.

Fall 3. Falls $r_{av} = \vartheta^2 \cdot r_{au}$ und $r_{bu} = r_{uv} = r_{bv} = \vartheta r_{au}$ ist, geht (9) über in

$$r_{ab} = \vartheta r_{au}^2 \frac{1 + \vartheta^2}{1 + \vartheta r_{au}}, \quad (15)$$

und es wird $|r_{ab}| < |r_{au}|$, wenn

$$\vartheta^3 < \frac{1}{|r_{au}|}. \quad (16)$$

Sofern wieder r_{au} der absolut grösste der Korrelationskoeffizienten ist, die einen direkten Zusammenhang ausdrücken, und ϑ zwischen 0 und 1 liegt, ist Ungleichung (16) unter den Bedingungen des Falles 3 stets erfüllt. Es ist somit auch $|r_{ab}| < |r_{au}|$.

Fall 4. Falls $r_{uv} = \vartheta^2 \cdot r_{au}$ und $r_{bu} = r_{av} = r_{bv} = \vartheta r_{au}$, geht (9) über in

$$r_{ab} = \vartheta r_{au}^2 \frac{1 + \vartheta}{1 + \vartheta^2 \cdot r_{au}}, \quad (17)$$

und es wird $|r_{ab}| < |r_{au}|$, wenn

$$\vartheta |r_{au}| < 1. \quad (18)$$

Sofern für r_{au} und ϑ dasselbe wie für die vorher erörterten Fälle gilt, ist Ungleichung (18) stets erfüllt, und damit ist auch hier stets $|r_{ab}| < |r_{au}|$.

Fall 5. Falls $r_{av} = r_{bu} = \vartheta r_{au}$ und $r_{uv} = r_{bv} = \lambda \vartheta r_{au}$, geht (9) über in

$$r_{ab} = \vartheta r_{au} + \frac{\lambda \vartheta^2 r_{au} (1 - \lambda r_{au}) (1 - \vartheta r_{au})}{1 - \lambda^2 \vartheta^2 r_{au}^2}. \quad (19)$$

und es wird $|r_{ab}| < |r_{au}|$, wenn

$$\lambda \vartheta^2 |r_{au}| < 1. \quad (20)$$

Sofern wieder r_{au} der absolut grösste der Korrelationskoeffizienten ist, welche einen direkten Zusammenhang ausdrücken, und sowohl λ wie auch ϑ zwischen Null und $+1$ liegen, ist Ungleichung (20) stets erfüllt. Hiermit ist unter den Bedingungen des Falles 5 auch stets $|r_{ab}| < |r_{au}|$.

Fall 6. Falls $r_{bv} = \lambda r_{uv}$, geht (9) über in

$$r_{ab} = r_{au} \cdot r_{bu} + \frac{(r_{av} - r_{au} \cdot r_{uv}) (\lambda - r_{bu}) r_{uv}}{1 - r_{uv}^2}, \quad (21)$$

und bei positiven Korrelationskoeffizienten wird $r_{ab} < r_{au}$, wenn

$$(r_{av} - r_{au} \cdot r_{uv}) (\lambda - r_{bu}) r_{uv} < r_{au} (1 - r_{bu}) (1 - r_{uv}^2). \quad (22)$$

Diese Ungleichung ist für $\lambda = 1$ erfüllt, wenn $r_{av} \cdot r_{uv} < r_{au}$, und sie ist für $\lambda = 0$ erfüllt, wenn $-r_{bu} (r_{au} - r_{av} \cdot r_{uv}) < r_{au} (1 - r_{uv}^2)$. Sofern wieder r_{au} der grösste der Korrelationskoeffizienten der direkten Zusammenhänge ist, gelten die letztgenannten beiden Ungleichungen, und es gilt Ungleichung (22) infolge ihrer Linearität in λ auch für alle λ zwischen 0 und $+1$. Hiermit gilt auch unter den Bedingungen des Falles 6, dass $r_{ab} < r_{au}$.

Auf Grund der aufgeführten hinreichenden Bedingungen scheint der maximal starke Zusammenhang immer dann ein direkter Zusammenhang zu sein, wenn im Feld der Korrelationskoeffizienten, welche direkte Zusammenhänge ausdrücken, gewisse Gleichmässigkeiten herrschen. Bei extravagantem Abweichen von diesen Gleichmässigkeiten (wenn z.B. wie im Fall $r_{uv} = 0$, $r_{au} = 0,70$ und $r_{bu} = r_{av} = r_{bv} = 0,60$ ein extravagantes Minimum im Feld der Korrelationskoeffizienten vorliegt) braucht der maximal starke Zusammenhang kein direkter Zusammenhang zu sein.

III. Das Kriterium verschwindender partieller Korrelationskoeffizienten. Es kann zwar praktisch damit gerechnet werden, dass die aufgezeigten Gleichmässigkeiten im Feld der Korrelationskoeffizienten direkter Zusammenhänge häufig vorkommen oder ungefähr vorliegen, doch besteht natürlich keine Sicherheit hierfür. Die Auffindung des stärksten Zusammenhangs innerhalb einer Vielheit von Beobachtungsobjekten berechtigt also noch nicht zu der Annahme, einen direkten Zusammenhang gefunden zu haben.

Ganz abgesehen hiervon ist es ausserdem ja auch ungewiss, ob sich in der statistisch erfassten Vielheit von Beobachtungsobjekten überhaupt ein direkter Zusammenhang befindet.

Auf jeden Fall ist es nötig, durch Ausschaltung von Beobachtungsobjekten schliesslich zu einem partiellen Korrelationskoeffizienten mit dem Wert Null zu gelangen. Bei der Auswahl der auszuschaltenden Beobachtungsobjekte fährt man wahrscheinlich am zweckmässigsten, indem man mit der Ausschaltung des stärksten Zusammenhangs beginnt, dann den zweitstärksten hinzunimmt und die Reihe der Ausschaltungen auf diese Art fortsetzt, bis man entweder zu einem partiellen Korrelationskoeffizienten mit dem Wert Null kommt oder zu dem Ergebnis gelangt, dass die Reihe der partiellen Korrelationskoeffizienten im Gefolge der hinzukommenden Ausschaltungen gegen einen von Null verschiedenen Wert geht, der nach (5) zwischen dem Anfangswert der Reihe und dem Wert Null liegt.

Es ist dabei natürlich möglich, in die Reihe der ausgeschalteten Beobachtungsobjekte nicht bloss solche Objekte hineinzubekommen, die einen direkten Zusammenhang mit dem ins Auge gefassten Untersuchungsgegenstand aufweisen, sondern es können sich in der Reihe der ausgeschalteten Objekte auch solche finden, die einen indirekten Zusammenhang mit dem Untersuchungsobjekt besitzen. Wenn jedoch die Reihe der zugehörigen partiellen Korrelationskoeffizienten (bei steigender Ordnung) den Wert Null erreicht, dann befinden sich unter den ausgeschalteten Beobachtungsobjekten gewiss alle diejenigen, die einen direkten Zusammenhang mit dem Untersuchungsobjekt haben.

Die Trennung der Beobachtungsobjekte, die einen direkten Zusammenhang mit dem Untersuchungsobjekt haben, von den in der Ausschaltungsreihe eventuell auch vorhandenen Beobach-

tungsobjekten, die einen indirekten Zusammenhang mit dem Untersuchungsgegenstand haben, ist eine weitere Aufgabe, die in Teil 2 der Arbeit behandelt wird.

C. Ein praktisches Beispiel

I. Das statistische Material. An Hand der Statistik der Vereinten Nationen möge untersucht werden, ob es auf dem beschriebenen Wege gelingt, in der Krebsforschung voranzukommen. Wie in der Arbeit Mittmann: *Über einen Zusammenhang zwischen Krebssterblichkeit und Ernährung*, « Krebsarzt » 2, 12 (1957) wird das Material von 29 Ländern herangezogen, die als statisch hinreichend zuverlässig angesehen werden dürfen. Neben dem Untersuchungsgegenstand Krebssterblichkeit werden alle diejenigen Beobachtungsobjekte der UN-Statistik herangezogen, die einen möglichst starken Zusammenhang mit der Krebssterblichkeit aufweisen und dabei auch in der Statistik sämtlicher 29 Länder aufgeführt werden. Es sind die folgenden Objekte aus dem Beobachtungsjahr 1952 und pro Kopf der Bevölkerung des jeweiligen Landes: Verbrauch tierischer Fettkalorien, Verbrauch tierischer Proteinkalorien, das Durchschnittsalter, Verbrauch an Kilowattstunden, Energie-Verbrauch, Kreislaufsterblichkeit, Zahl der Telefonteilnehmer, Zementverbrauch, Kalorien tierischer Kohlehydrate, Zahl der Radio-Empfänger, Verbrauch an raffiniertem Zucker, Zahl der Tageszeitungen, Zahl der beförderten Briefe, Kalorien vegetarischen Ursprungs, die Säuglings-Überlebenswahrscheinlichkeit, die Einkommenshöhe, Zahl der Ärzte, der Export, die Zahl der Personenkraftwagen, der Import und die Diabetes-Sterblichkeit.

II. Die berechneten Korrelationskoeffizienten. Die gewöhnlichen Korrelationskoeffizienten und die partiellen Korrelationskoeffizienten erster Ordnung mögen in Tabellenform vor Augen geführt werden.

Tabelle 1 zeigt die Werte der gewöhnlichen Korrelationskoeffizienten zwischen allen genannten Beobachtungsobjekten (231 Werte). In der ersten Zeile z.B. findet man im ersten Feld links den gewöhnlichen Korrelationskoeffizienten zwischen der Krebssterblichkeit und den tierischen Fettkalorien, das zweite Feld

der ersten Zeile gibt den Korrelationskoeffizienten zwischen der Krebssterblichkeit und den tierischen Proteinkalorien, usw.; der Durchschnittswert der in der ersten Zeile enthaltenen gewöhnlichen Korrelationskoeffizienten beträgt 0,720. Das erste Feld der zweiten Zeile ist leer, das zweite Feld der zweiten Zeile gibt den Korrelationskoeffizienten zwischen den tierischen Fettkalorien und den tierischen Proteinkalorien, das dritte Feld der zweiten Zeile gibt den Korrelationskoeffizienten zwischen den tierischen Fettkalorien und dem Durchschnittsalter, usw., und so fort.

Tabelle 2 zeigt alle in Frage kommenden partiellen Korrelationskoeffizienten erster Ordnung zwischen der Krebssterblichkeit und je einem der andern genannten Beobachtungsobjekte unter Ausschaltung des Einflusses je eines der übrigen Beobachtungsobjekte (420 Werte).

Diese Tabelle ist so angelegt, dass in einer Zeile alle partiellen Korrelationskoeffizienten zwischen der Krebssterblichkeit und ein und demselben sonstigen Beobachtungsobjekt stehen, wobei ein Unterschied jeweils nur durch das ausgeschaltete Objekt entsteht. So stehen z.B. in der ersten Zeile alle partiellen Korrelationskoeffizienten zwischen der Krebssterblichkeit und den tierischen Fettkalorien, wobei links das erste Feld leer ist, das zweite Feld von links gibt diesen partiellen Korrelationskoeffizienten unter Ausschaltung der tierischen Proteinkalorien wieder, das dritte Feld gibt diesen partiellen Korrelationskoeffizienten unter Ausschaltung des Durchschnittsalters, usw. Am rechten Rand der jeweiligen Zeile ist vermerkt, wie gross der Durchschnittswert in der Zeile ist. Am rechten Rand der ersten Zeile von oben steht z.B. der Durchschnittswert 0,735.

Demgegenüber enthält eine Spalte der Tabelle 2 alle partiellen Korrelationskoeffizienten mit dem gleichen ausgeschalteten Objekt. So stehen z.B. in der ersten Spalte von links alle partiellen Korrelationskoeffizienten zwischen der Krebssterblichkeit und je einem der andern Objekte unter Ausschaltung der tierischen Fettkalorien, wobei oben das erste Feld leer ist, das zweite Feld von oben gibt den partiellen Korrelationskoeffizienten zwischen der Krebssterblichkeit und den tierischen Proteinkalorien unter Ausschaltung der tierischen Fettkalorien, das dritte Feld gibt den partiellen Korrelationskoeffizienten zwischen der Krebssterb-

lichkeit und dem Durchschnittsalter unter Ausschaltung der tierischen Fettkalorien usw. Am unteren Rande der jeweiligen Spalte ist vermerkt, wie gross der Durchschnittswert in der betreffenden Spalte ist. Am unteren Rande der ersten Spalte von links steht z.B. der Durchschnittswert 0,187.

Macht man die am rechten Rand der Tabelle 2 stehenden Durchschnittszahlen zur Grundlage einer Ausschaltungs-Anordnung (was aus Selektionsgründen ratsamer als eine Zugrundelegung der gewöhnlichen Korrelationskoeffizienten ist), so lautet die Reihe der auszuschaltenden Objekte: Tierische Fettkalorien, tierische Protein-Kalorien, Durchschnittsalter, Kilowattstunden-Verbrauch, usw. Am Ende dieser Reihe befinden sich: Export, Zahl der Personenkraftwagen, Import und die Diabetes-Sterblichkeit als Schlussglied.

Nun zur Berechnung der partiellen Korrelationskoeffizienten höherer Ordnung! Eine ebenso ausführliche Berechnung der partiellen Korrelationskoeffizienten höherer Ordnung, wie sie bei den partiellen Korrelationskoeffizienten erster Ordnung arbeitsmässig geleistet werden kann, ist bei den partiellen Korrelationskoeffizienten höherer Ordnung aus Zeitgründen kaum durchführbar. Nach erfolgter Festlegung der Ausschaltungs-Anordnung begnügen wir uns also damit, die partiellen Korrelationskoeffizienten höherer Ordnung wohl zwischen der Krebssterblichkeit und jedem andern Beobachtungsobjekt, aber nur unter Ausschaltung der *Anfangsglieder* der Ausschaltungsreihe zu berechnen.

Die Berechnung der partiellen Korrelationskoeffizienten zweiter Ordnung erfolgt nach Formel (7). Hierbei können die notwendigen partiellen Korrelationskoeffizienten erster Ordnung teilweise aus Tabelle 2 entnommen werden, wogegen andere erst mit Hilfe der gewöhnlichen Korrelationskoeffizienten der Tabelle 1 berechnet werden müssen.

Die Berechnung der partiellen Korrelationskoeffizienten dritter Ordnung erfolgt auf dem Weg über partielle Korrelationskoeffizienten zweiter Ordnung nach der Formel

$$r_{ab,uvw} = \frac{r_{ab,uv} - r_{aw,uv} r_{bw,uv}}{\sqrt{(1 - r_{aw,uv}^2)(1 - r_{bw,uv}^2)}} \quad (23)$$

wobei ein drittes zur Ausschaltung kommendes Objekt mit dem Buchstaben w bezeichnet ist.

Die Berechnung der partiellen Korrelationskoeffizienten vierter Ordnung erfolgt mit Hilfe von partiellen Korrelationskoeffizienten dritter Ordnung gemäss der Formel

$$r_{ab,uvw\alpha} = \frac{r_{ab,uvw} - r_{a\alpha,uvw} r_{b\alpha,uvw}}{\sqrt{(1 - r_{a\alpha,uvw}^2)(1 - r_{b\alpha,uvw}^2)}}, \quad (24)$$

wobei der Index α das vierte zur Ausschaltung gelangende Objekt bezeichnet.

Bei einem Vergleich der Formeln (7), (23) und (24) sieht man ohne weiteres, nach welcher Rekursionsformel die Berechnung partieller Korrelationskoeffizienten von noch höherer Ordnung zu erfolgen hätte.

Bei Ausschaltung der tierischen Fettkalorien und der tierischen Proteinkalorien ergibt sich als partieller Korrelationskoeffizient zwischen der Krebssterblichkeit und je einem der übrigen 19 Beobachtungsobjekte im Durchschnitt der Wert 0,109. Bei Ausschaltung der tierischen Fettkalorien, der tierischen Proteinkalorien und des Durchschnittsalters ergibt sich als partieller Korrelationskoeffizient zwischen der Krebssterblichkeit und je einem der übrigen 18 Beobachtungsobjekte durchschnittlich der Wert 0,054. Bei Ausschaltung der tierischen Fettkalorien, der tierischen Proteinkalorien, des Durchschnittsalters und der Kilowattstunden-Zahl ergibt sich als partieller Korrelationskoeffizient zwischen der Krebssterblichkeit und je einem der übrigen 18 Beobachtungsobjekte durchschnittlich der Wert $-0,038$.

Abb. 1 führt alle soeben genannten Werte noch einmal in Gestalt der ausgezogenen Kurve vor Augen. Als Anfangswert der Reihe wird dabei der Durchschnittswert aller 21 gewöhnlichen Korrelationskoeffizienten zwischen der Krebssterblichkeit und je einem der 21 anderen Objekte angeführt; es ist dies der Wert 0,720. Wie zu sehen ist, geht die Reihe der partiellen Korrelationskoeffizienten gegen den Wert Null. Der partielle Korrelationskoeffizient vierter Ordnung geht sogar leicht ins Negative. Sofern der Befund nicht auf Zufalls-Schwankungen zurückzuführen ist, gelangen wir also zu dem Ergebnis: Die 4 ausgeschalteten Objekte, nämlich die tierischen Fett-Kalorien, die tierischen Protein-Kalorien, das Durchschnitts-Alter und die Kilowattstunden-Zahl, weisen Zusammenhänge mit der Krebssterblichkeit

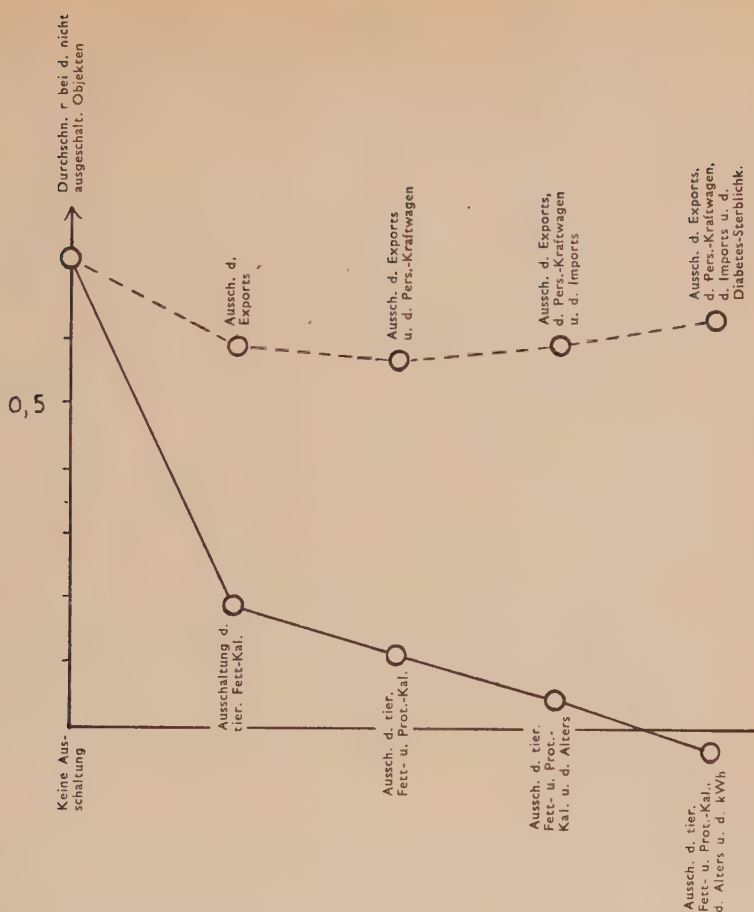


Abb. 1: Der Korrelationskoeffizient r , der durchschnittlich zwischen den nicht ausgeschalteten Beobachtungsobjekten und der Krebssterblichkeit besteht, falls

- a) zuerst die tierischen Fettkalorien, dann zusätzlich die tierischen Proteinkalorien, dann zusätzlich das Durchschnittsalter und dann noch zusätzlich die Kilowattstunden ausgeschaltet werden (Ausgezogene Kurve),
- b) zuerst der Export, dann zusätzlich die Personenkraftwagen, dann zusätzlich der Import und dann noch zusätzlich die Diabetes-Sterblichkeit ausgeschaltet werden (Gestrichelte Kurve).

auf, unter denen sich die direkten Zusammenhänge befinden; dabei könnte es auch sein, dass *jedes* dieser 4 Objekte einen direkten Zusammenhang mit der Krebssterblichkeit besitzt. Voraussetzung für die Gültigkeit des Ergebnisses ist natürlich seine statistische Sicherheit.

III. Die statistische Sicherheit des Ergebnisses. Eine Prüfung der statistischen Sicherheit des Ergebnisses ist auf empirischem und theoretischem Wege durchführbar. Der empirische Weg ist gewichtiger, weil der theoretische die etwas fragwürdige Unabhängigkeits-Voraussetzung benutzt.

Auf empirischem Wege lässt sich die Prüfung durchführen, indem man vergleichsweise nicht diejenigen Beobachtungs-Objekte ausschaltet, die die relativ stärksten Zusammenhänge mit der Krebssterblichkeit haben, sondern diejenigen Beobachtungs-Objekte, die die vier relativ schwächsten Zusammenhänge mit der Krebssterblichkeit aufweisen, nämlich Export, Zahl der Personen-Kraftwagen, Import und Diabetes-Sterblichkeit. Abbildung 1 zeigt das Ergebnis in Gestalt der gestrichelten Kurve. An Hand der gestrichelten Kurve, die offenbar nicht gegen Null geht, ist ersichtlich, in welchem Rahmen sich etwa die Zufalls-Schwankungen bewegen können. Es erscheint hiernach unwahrscheinlich, dass die ausgezogene Kurve nur infolge von Zufalls-Schwankungen den Wert Null erreicht.

Auf theoretischem Wege kann die statistische Sicherheit des Ergebnisses dadurch geprüft werden, dass man den mittleren Fehler des partiellen Korrelationskoeffizienten r für die Hypothese $r = 0$ berechnet und den Wert von r mit dem mittleren Fehler $\sigma(r)$ vergleicht. Der Quotient aus r und dem mittleren Fehler $\sigma(r)$ ist dann ein Mass für die Güte der Ubereinstimmung von r mit dem hypothetischen Wert $r = 0$. Abbildung 2 gibt eine entsprechende Darstellung für diesen Quotienten, wie sie durch Abbildung 1 für r gegeben wird. Bei Annahme einer Gauss-schen Fehler-Verteilung (Normal-Verteilung) braucht der Quotient nur kleiner als $+3$ zu sein, um des Schluss zuzulassen, dass sich der Befund nicht in Widerspruch zur Hypothese $r = 0$ befindet. Während nun die gestrichelte Kurve der Abbildung 2 etwa gegen den Ordinaten-Wert 14 geht und hier also von einer Vereinbarkeit mit der Hypothese $r = 0$ überhaupt keine Rede sein kann,

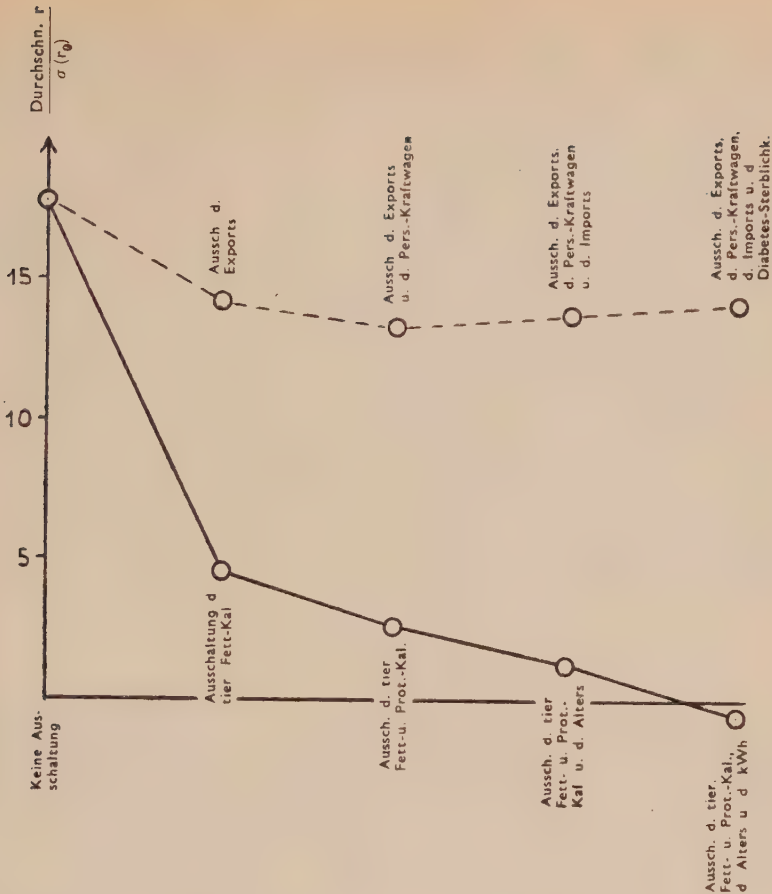


Abb. 2 : Der Quotient aus dem Korrelationskoeffizienten r , der zwischen der Krebssterblichkeit und den nicht ausgeschalteten Beobachtungsobjekten durchschnittlich besteht, und dem mittleren Fehler $\sigma(r)$, der für die Hypothese $r = 0$ berechnet wird, falls

- zuerst die tierischen Fettkalorien, dann zusätzlich die tierischen Proteinkalorien, dann zusätzlich das Durchschnittsalter und dann noch zusätzlich die Kilowattstunden ausgeschaltet werden (Ausgezogene Kurve),
- zuerst der Export, dann zusätzlich die Personenkraftwagen, dann zusätzlich der Import und dann noch zusätzlich die Diabetes-Sterblichkeit ausgeschaltet werden (Gestrichelte Kurve)

durchschneidet die ausgezogene Kurve der Abbildung 2 den Ordinaten-Wert $+3$ bereits beim Übergang vom partiellen Korrelationskoeffizienten erster Ordnung zum partiellen Korrelationskoeffizienten zweiter Ordnung, d.h. nach Ausschaltung sowohl der tierischen Fett-Kalorien wie auch der tierischen Protein-Kalorien; der genannte Quotient zeigt den Wert $+2,55$. Nach zusätzlicher Ausschaltung des Durchschnitts-Alters nimmt der Quotient den Wert $+1,23$ an. Bei zusätzlicher Ausschaltung des Kilowattstunden-Verbrauchs durchschneidet die ausgezogene Kurve der Abbildung 2 den Wert Null, und der Quotient der Übereinstimmungs-Güte nimmt den Wert $-0,84$ an. Das letzte bedeutet unter der Voraussetzung einer Gauss-schen Fehler-Verteilung, dass nur eine Wahrscheinlichkeit von 20 % für einen partiellen Korrelationskoeffizienten im Bereich positiver Werte bestünde. Eine solche Aussage bedeutet natürlich mehr-als die Aussage, dass sich der Befund nicht in Widerspruch zur Hypothese $r = 0$ befindet.

TEIL 2

In Teil 1 der Arbeit wurde ein Verfahren aufgezeigt, mit dem man die Frage entscheiden kann, ob in einem statistischen Material im Hinblick auf einen bestimmten Untersuchungs-Gegenstand sämtliche praktisch wirksamen Ursachen enthalten sind oder nicht; man kann bejahendenfalls eine Teil-Gesamtheit von Beobachtungs-Gegenständen ausfindig machen, innerhalb derer die gesuchten Ursachen enthalten sind. Die Frage, wie man anschliessend zu beurteilen vermag, ob *alle* dieser Teil-Gesamtheit angehörenden Gegenstände Ursachen sind oder welche von ihnen die gesuchten Ursachen sind, soll in Teil 2 behandelt werden.

Ausserdem soll noch ein anderer Weg vor Augen geführt werden, auf dem man die gesuchten Ursachen direkt ausfindig machen kann. Aus Kontroll-Gründen ist es zweckmässig, beide Wege zu gehen.

D. Verfahren der nachträglichen Ursachen-Bestimmung

Das Verfahren der nachträglichen Ursachen-Bestimmung, das an Teil 1 der Arbeit anknüpft, benutzt das Verschwinden des partiellen Korrelationskoeffizienten für den Fall, dass sämt-

liche wirksamen Ursachen ausgeschaltet werden. Ein Verschwinden erfolgt dann und nur dann, wenn sämtliche wirksamen Ursachen ausgeschaltet werden. Auf Grund der entwickelten Methode ist es dabei nun möglich, dass sich unter den Grössen, deren Ausschaltung zu einem verschwindenden Korrelationskoeffizienten führt, auch Grössen befinden, die keinen ursächlichen, sondern einen indirekten Zusammenhang mit dem Untersuchungsgegenstand haben. Es ist erforderlich zu klären, welche Grössen bei der Ausschaltung eventuell fortgelassen werden können, ohne das erwünschte Verschwinden des Korrelationskoeffizienten zu beeinträchtigen. Hierzu müssen die verschiedenen Möglichkeiten der Ausschaltung durchforscht werden.

I. Die verschiedenen Möglichkeiten der Ausschaltung von α aus n Grössen. Ist n die Anzahl der Grössen, deren Ausschaltung zu dem Korrelationskoeffizienten Null geführt hat, und wollen wir nicht alle n Grössen, sondern nur α von ihnen ausschalten, so ist

$$\binom{n}{\alpha} = \frac{n(n-1)\dots(n-\alpha+1)}{1 \cdot 2 \cdot \dots \cdot \alpha} \quad (25)$$

die zugehörige Zahl der sog. « Kombinationen ohne Wiederholung ». Ist z.B. $n=4$ und seien a, b, c und d die n Grössen, dann gibt es $\binom{4}{1} = 4$ Möglichkeiten dafür, dass nur eine von

ihnen für die Ausschaltung benutzt wird. Es gibt $\binom{4}{2} = \frac{4 \cdot 3}{1 \cdot 2} = 6$ Möglichkeiten dafür, dass je 2 der 4 Grössen für die Ausschaltung benutzt werden, nämlich die Möglichkeiten ab, ac, ad, bc, bd und cd . Es gibt $\binom{4}{3} = \frac{4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 3} = 4$ Möglichkeiten dafür, dass je 3 der 4 Grössen für die Ausschaltung benutzt werden, nämlich die Möglichkeiten abc, abd, acd und bcd .

Da für α alle Werte von 1 bis $n-1$ in Frage kommen, beträgt die Gesamtzahl aller Möglichkeiten, dass weniger als n Grössen für die Ausschaltung benutzt werden.

$$\binom{n}{1} + \binom{n}{2} + \dots + \binom{n}{n-1} = \sum_{\alpha=1}^{n-1} \binom{n}{\alpha} = 2^n - 2. \quad (26)$$

Für $n = 4$ wären es also $2^4 - 2 = 14$ Möglichkeiten, die untersucht werden müssen, um eine Entscheidung darüber treffen zu können, wieviele und welche der 4 Grössen für die Ausschaltung nötig sind, damit der Korrelationskoeffizient gleich Null wird.

II. Zur Durchführung der erforderlichen Rechnungen. Die Zahl der partiellen Korrelationskoeffizienten, die man zusätzlich zu den im ersten Teil der Arbeit berechneten benötigt, ist natürlich kleiner als $2^n - 2$, da man ja einerseits schon alle partiellen Korrelationskoeffizienten erster Ordnung und andererseits je einen der partiellen Korrelationskoeffizienten höherer Ordnung berechnet hat, also $n + (n - 2) = 2n - 2$ partielle Korrelationskoeffizienten bereits kennt. Die Anzahl der noch zu bestimmenden partiellen Korrelationskoeffizienten beträgt also bloss $2^n - 2 - 2n + 2 = 2^n - 2n$, im Falle $n = 4$ z.B. $2^4 - 8 = 8$.

Die Formeln, die man zur Durchführung der Rechnungen braucht, sind im ersten Teil der Arbeit aufgeführt. Zur Bestimmung eines partiellen Korrelationskoeffizienten unter Ausschaltung einer Grösse benötigt man Formel (1), bei Ausschaltung von 2 Grössen braucht man Formel (7), bei Ausschaltung von 3 Grössen die Formel (23), bei Ausschaltung von 4 Grössen die Formel (24), bei Ausschaltung von mehr als 4 Grössen eine entsprechend gebaute Formel. Dabei kann man die partiellen Korrelationskoeffizienten niederer Ordnung stets bei der Berechnung der partiellen Korrelationskoeffizienten höherer Ordnung verwenden.

III. Praktisches Beispiel. Das im ersten Teil der Arbeit behandelte Beispiel möge hier weiter behandelt werden. Es wurde gefunden, dass die Ausschaltung der tierischen Fettkalorien, der tierischen Proteinkalorien, des Durchschnittsalters und der Kilowattstundenzahl zu einem durchschnittlichen Verschwinden der Korrelationskoeffizienten zwischen der Krebssterblichkeit und je einem der übrigen Beobachtungsgegenstände führt. Es fragt sich, ob eine Ausschaltung der tierischen Fettkalorien sowie der tierischen Proteinkalorien sowie des Durchschnittsalters und der Kilowattstundenzahl notwendig ist, um

zu einem durchschnittlich verschwindenden Korrelationskoeffizienten zu gelangen oder ob die Ausschaltung in ihrem Ausmass verringert werden kann.

Da die Anzahl der ausgeschalteten Beobachtungsgegenstände 4 beträgt, gibt es 14 Möglichkeiten der Ausschaltung von weniger als 4 Objekten. Dies sind, wenn wir die 4 Objekte mit u , v , w und x bezeichnen, die folgenden Möglichkeiten:

$$u, v, w, x,$$

$$uv, uw, ux, vw, vx, wx,$$

$$uvw, uvx, uwx \text{ und } vwx.$$

Hiervon sind die Möglichkeiten u, v, w, x, uv und uvw bereits im ersten Teil der Arbeit behandelt worden. Es bleiben die restlichen 8 zu untersuchen.

Die Ergebnisse seien wieder für den Quotienten $\frac{\text{Durchschn. } r}{\sigma(r_0)}$ vor Augen geführt, wobei das durchschnittliche r derjenige Korrelationskoeffizient ist, der zwischen der Krebssterblichkeit und den nicht ausgeschalteten Beobachtungsobjekten durchschnittlich gefunden wird, und $\sigma(r_0)$ der zugehörige mittlere Fehler bei Zugrundelegung der Hypothese $r = 0$ ist.

Tabelle 3 zeigt die Ergebnisse im einzelnen. Zwecks Vereinfachung der Wiedergabe sind dabei die tierischen Fett-Kalorien mit tF , die tierischen Protein-Kalorien mit tP , das Durchschnittsalter mit A und die Kilowattstunden-Zahl mit kW bezeichnet. Man sieht, dass die Ausschaltung von einem der 4 Objekte nur zu positiven Quotienten-Werten führt. Das gleiche gilt für die Fälle der Ausschaltung von je 2 der 4 Objekte. Dagegen treten bei Ausschaltung von 3 Objekten schon negative Quotienten-Werte auf; die Ausschaltungskombination Tierische Fett-Kalorien/Durchschnittsalter/Kilowattstunden-Zahl weist dabei mit $-0,29$ den stärksten negativen Quotientenwert auf und besitzt somit die *grösste Wahrscheinlichkeit dafür, dass der Korrelationskoeffizient r nicht grösser als Null ist.*

Betrachten wir die Tabellenzahlen hinsichtlich der statistischen Vereinbarkeit des Ergebnisses mit der Hypothese $r = 0$, d.h. im Hinblick auf Quotient-Werte unterhalb 3, dann sehen wir, dass solche Vereinbarkeiten für die Ausschaltungskombinationen

TABELLE 3: Der Quotient aus dem Korrelationskoeffizienten r , der zwischen der Krebssterblichkeit und jedem der nicht ausgeschalteten Beobachtungs-Objekte durchschnittlich besteht, und dem zugehörigen mittleren Fehler, der für die Hypothese $r = 0$ berechnet wird, falls die tierischen Fettkalorien, die tierischen Proteinkalorien, das Durchschnittsalter und die Kilowattstunden-Zahl in den verschiedensten Kombinationen ausgeschaltet werden. Gebrauchte Abkürzungen: tF = tierische Fett-Kalorien, tP = tierische Protein-Kalorien, A = Durchschnittsalter, kW = Kilowattstunden-Zahl.

Keine Ausschaltung: + 17,76

Ausschaltung von	tF	tP	A	kW			
	+ 4,49	+ 3,37	+ 11,78	+ 8,07			
Ausschaltung von	tF	tF	tF	tP	tP	A	
	tP	A	kW	A	kW	kW	
	+ 2,55	+ 3,78	+ 0,03	+ 1,45	+ 1,61	+ 5,34	
Ausschaltung von	tF	tF	tF	tP			
	tP	tP	A	A			
	A	kW	kW	kW			
	+ 1,23	— 0,17	— 0,29	— 0,12			
Ausschaltung von	tF						
	tP						
	A						
	kW						
	— 0,84						

tF , tP , kW , A und kW auftreten, wobei die Ausschaltungskombination Tierische Fettkalorien/Kilowattstundenzahl mit dem Quotientenwert +0,03 am besten mit dem Wert Null vereinbar ist. — Die Hypothesen-Trennschärfe lässt, wie man sieht, sehr zu wünschen übrig.

IV. Ein fingiertes Beispiel. Das vorstehend behandelte praktische Beispiel stellt einen Normalfall dar, der weiter keine Tücken enthält. Es interessiert natürlich, wie die angewandte Methode funktionieren würde, wenn der aussergewöhnliche Fall vorläge, dass sich im Feld der Korrelationskoeffizienten, welche die direkten Zusammenhänge darstellen, ein extravagantes Minimum befindet, sodass der stärkste auftretende Zusammenhang

nicht ein kausaler, sondern ein indirekter Zusammenhang ist. Es sei hierzu an Formel (10) angeknüpft, die einen solchen Fall zum Gegenstand hat. Es dreht sich hier um den Fall, dass 2 Variable a und b indirekt über 2 Ursachen u und v miteinander zusammenhängen und dabei $r_{uv} = 0$ ist. Wie Formel (10) besagt, ist dann $r_{ab} = r_{au} \cdot r_{bu} + r_{av} \cdot r_{bv}$. Mit Hilfe dieser Formel kann man leicht ein Beispiel der gewünschten Art konstruieren, indem z.B. $r_{au} = 0,70$ und $r_{av} = 0,60$ gewählt wird, während $r_{bu} = r_{bv} = 0,55$ ist; dann wird $r_{ab} = 1,30 \cdot 0,55 = 0,715$, und wir hätten damit den gewöhnlichen Korrelationskoeffizienten zwischen einem Untersuchungsobjekt a und einem Beobachtungsobjekt b . Bei Austausch des Objektes b mit einem Objekt c , für welches $r_{cu} = r_{cv} = 0,50$ sein möge, wird $r_{ac} = 1,30 \cdot 0,50 = 0,650$. Bei Austausch von b mit einem Objekt d , für welches $r_{du} = r_{dv} = 0,45$ sein möge, wird $r_{ad} = 1,30 \cdot 0,45 = 0,585$. Bei Austausch von b mit einem Objekt e , für welches $r_{eu} = r_{ev} = 0,40$ sein möge, wird $r_{ae} = 1,30 \cdot 0,40 = 0,520$. Hiermit hätten wir auch die gewöhnlichen Korrelationskoeffizienten zwischen a und weiteren Beobachtungsobjekten c , d und e . Auf gleiche Weise erhalten wir die gewöhnlichen Korrelationskoeffizienten zwischen den Beobachtungsobjekten b , c , d und e . Es wird z.B. $r_{bc} = 2 \cdot 0,55 \cdot 0,50 = 0,550$, ferner $r_{bd} = 2 \cdot 0,55 \cdot 0,45 = 0,495$, usw. Tabelle 4 zeigt alle diese Werte. Ausserdem enthält sie die Werte der Korrelationskoeffizienten, die die Zusammenhänge mit den Ursachen u und v ausdrücken. Die Ursachen u und v erfahren bei der anschliessenden Erprobung der Methode eine Behandlung als Beobachtungsobjekte.

Der höchste der vorkommenden Korrelationskoeffizienten, nämlich $r_{ab} = 0,715$, stellt einen indirekten Zusammenhang dar. Die ursächlichen Zusammenhänge, die im Beispiel vorkommen, folgen erst später. Der stärkste Zusammenhang mit u ist durch den Korrelationskoeffizienten $r_{au} = 0,700$ gegeben, und der stärkste Zusammenhang mit v ist durch den Korrelationskoeffizienten $r_{av} = 0,600$ gegeben. Der durchschnittliche Korrelationskoeffizient zwischen dem Untersuchungsobjekt a und je einem der Beobachtungsobjekte b bzw. c bzw. d bzw. e bzw. u bzw. v beträgt $3,770 : 6 = 0,628$.

Die nach Formel (1) zu findenden partiellen Korrelationskoeffizienten erster Ordnung zwischen a und je einem der Be-

TABELLE 4: Die gewöhnlichen Korrelationskoeffizienten im fingierten Beispiel mit dem Untersuchungsobjekt *a* und den Beobachtungsobjekten *b*, *c*, *e*, *u* und *v*.

	b	c	d	e	u	v
a	0,715	0,650	0,585	0,520	0,700	0,600
b		0,550	0,495	0,440	0,550	0,550
c			0,450	0,400	0,500	0,500
d				0,360	0,450	0,450
e					0,400	0,400
u						0,000

beobachtungsobjekte unter Ausschaltung je eines der übrigen Beobachtungsobjekte ersieht man aus Tabelle 5. Diese Tabelle ist ebenso wie Tabelle 2 angelegt. Am linken Tabellenrand sind die Beobachtungsobjekte vermerkt, deren Zusammenhang mit dem Untersuchungsgegenstand *a* ausgedrückt wird, während am oberen Tabellenrand diejenigen Beobachtungsobjekte vermerkt sind, deren Ausschaltung erfolgt. Am rechten Tabellenrand ist auch wieder der Durchschnittswert der betreffenden Zeile vermerkt,

TABELLE 5: Der partielle Korrelationskoeffizient erster Ordnung zwischen dem Untersuchungsgegenstand *a* und den am linken Tabellenrand vermerkten Beobachtungsobjekten unter Ausschaltung der am oberen Tabellenrand vermerkten Beobachtungsobjekte

	b	c	d	e	u	v	
b		0,563	0,604	0,634	0,553	0,576	0,586
c	0,440		0,534	0,565	0,485	0,505	0,506
d	0,380	0,431		0,499	0,423	0,441	0,435
e	0,327	0,373	0,409		0,367	0,382	0,372
u	0,525	0,570	0,603	0,628		0,875	0,640
v	0,354	0,418	0,465	0,501	0,840		0,516
	0,405	0,471	0,523	0,565	0,534	0,556	

am unteren Tabellenrand der Durchschnittswert der betreffenden Spalte.

Das Maximum der am rechten Tabellenrand vermerkten Werte liegt mit 0,640 beim Beobachtungsobjekt u . Es folgt b mit dem Wert 0,586, v mit dem Wert 0,516, c mit 0,506, d mit 0,435 und e mit 0,372. Dies ist die Anordnung bei der nun folgenden Ausschaltung.

Bei Ausschaltung von u ergibt sich nach Tabelle 5 der durchschnittliche Korrelationskoeffizient 0,534. Bei Ausschaltung von u und b , die unter Benutzung von Formel (7) erfolgt, ergibt sich der durchschnittliche Korrelationskoeffizient 0,420. Bei Ausschaltung von u , b und v , die unter Benutzung von Formel (23) erfolgt, ergibt sich der durchschnittliche Korrelationskoeffizient 0,000. Da die Ausschaltung von u , b und v zu dem Korrelationskoeffizienten Null führt, sind in diesen 3 Objekten alle Ursachen enthalten.

Zur Prüfung der Frage, ob bereits die Ausschaltung einer Teilgesamtheit dieser 3 Objekte genügt, um den Korrelationskoeffizienten Null zu erhalten, sind die Ausschaltungen jeder einzelnen der 3 Grössen u , b und v und die Ausschaltungen jeder einzelnen der Kombinationen ub , uv und bv zu betrachten. Die Ergebnisse der ersten 3 genannten Ausschaltungen sind am unteren Rand der Tabelle 5 aufgeführt, die Ausschaltung von ub ist auch schon durchgeführt, und die Ausschaltungen von uv und bv sind noch durchzuführen. Die Ausschaltung von uv ergibt den durchschnittlichen Korrelationskoeffizienten 0,000, während die Ausschaltung von bv den durchschnittlichen Korrelationskoeff-

TABELLE 6: Der Korrelationskoeffizient, der zwischen a und jedem der nicht ausgeschalteten Beobachtungsobjekte durchschnittlich besteht, falls u , b und v in den verschiedensten Kombinationen ausgeschaltet werden.

Keine Ausschaltung:	0,628		
Ausschaltung von	u	b	v
	0,534	0,405	0,556
Ausschaltung von	ub	uv	bv
	0,420	0,000	0,446
Ausschaltung von	ubv		
	0,000		

fizienten 0,446 ergibt. Alle diese Werte sind in Tabelle 6 zusammengefasst.

Man sieht, dass der Korrelationskoeffizient Null bereits bei Ausschaltung von u und v erreicht wird, woraus folgt, dass b nur einen indirekten Zusammenhang mit a hat. Daneben sind die Korrelationskoeffizienten erster Ordnung sämtlich von Null verschieden, woraus folgt, dass u und v direkte Zusammenhänge mit a besitzen. Das Ergebnis stimmt genau mit der Konstruktion des fingierten Beispiels überein.

E. Verfahren der direkten Ursachen-Bestimmung

Die Grundlage des Verfahrens der nachträglichen Ursachenbestimmung besteht darin, dass die Ausschaltung der wirksamen Ursachen zum durchschnittlichen Korrelationskoeffizienten Null führt und dass umgekehrt das Verschwinden eines durchschnittlichen partiellen Korrelationskoeffizienten auf die Ausschaltung von Beobachtungsobjekten zurückgeht, die entweder alle wirksamen Ursachen zusammen mit anderen Beobachtungsobjekten enthalten oder, falls alle partiellen Korrelationskoeffizienten niedriger Ordnung nicht Null sind, die genau identisch mit den wirksamen Ursachen sind.

Das Verfahren der direkten Ursachenbestimmung benutzt den im Anschluss an die Ungleichungen (3) ausgesprochenen Satz: Der indirekte Zusammenhang zwischen 2 Beobachtungsobjekten a und b ist schwächer als der direkte Zusammenhang zwischen a und der (in diesem Satz vorausgesetzten alleinigen) Ursache u sowie der direkte Zusammenhang zwischen b und der Ursache u . Da sich zeigen lässt, dass dieser Satz auch für partielle Korrelationskoeffizienten Gültigkeit hat, ist er nicht nur für den Fall der Wirksamkeit einer einzigen Ursache, sondern ganz allgemein von Bedeutung.

I. Das Kriterium des maximal-starken Zusammenhangs. Aus Formel (1) folgt, dass $r_{ab,u} = 0$, wenn der Zähler in (1) verschwindet, d.h. wenn Formel (2) gilt: $r_{ab} - r_{au} \cdot r_{bu} = 0$. Aus Formel (7) folgt, dass $r_{ab,uv} = 0$, wenn (8) gilt:

$$r_{ab,u} - r_{av,u} \cdot r_{bv,u} = 0$$

oder

$$r_{ab,v} - r_{au,v} \cdot r_{bu,v} = 0,$$

was dasselbe ist.

Aus Formel (23) folgt, dass $r_{ab,uvw} = 0$, wenn

$$\left\{ \begin{array}{ll} r_{ab,uv} - r_{aw,uv} \cdot r_{bw,uv} = 0 & \text{oder anders geschrieben} \\ r_{ab,uw} - r_{av,uw} \cdot r_{bw,uw} = 0 & \text{oder anders ausgedrückt} \\ r_{ab,vw} - r_{au,vw} \cdot r_{bu,vw} = 0. \end{array} \right. \quad (27)$$

Entsprechendes gilt für partielle Korrelationskoeffizienten von höherer Ordnung.

α. Fall der alleinigen Wirksamkeit einer Ursache u. Wie gesagt, bedeutet die Gleichung $r_{ab} - r_{au} \cdot r_{bu} = 0$, dass $|r_{ab}|$ kleiner als $|r_{au}|$ ist, d.h.: Bei alleiniger Wirksamkeit der Ursache u , also bei verschwindendem $r_{ab,u}$, ist unter den Zusammenhängen die das Untersuchungsobjekt a mit andern Beobachtungsobjekten aufweist, der Zusammenhang mit u der maximal starke.

Falls bei Wirksamkeit der Ursache u eine Nicht-Ursache v ausgeschaltet wird, folgt aus (7) und (8) noch, dass $r_{ab,uv} = 0$ und $|r_{ab,v}| < |r_{au,v}|$, d.h.: Wird bei alleiniger Wirksamkeit von u nicht u , sondern ein anderes Beobachtungsobjekt v ausgeschaltet, so ist unter den Zusammenhängen, die das Untersuchungsobjekt a mit anderen Beobachtungsobjekten bei Ausschaltung von v aufweist, auch wieder der Zusammenhang mit u der maximal starke.

β. Fall der Wirksamkeit der beiden Ursachen u und v. Bei Wirksamkeit der beiden Ursachen u und v ist nach (9) keine Möglichkeit gegeben, den Zusammenhang, der durch den grössten *gewöhnlichen* Korrelationskoeffizienten ausgedrückt wird, ohne weiteres als direkten Zusammenhang anzusehen. Aus (8) folgt aber, dass die partiellen Korrelationskoeffizienten erster Ordnung im Falle zweier Ursachen dieselbe Rolle spielen wie die *gewöhnlichen* Korrelationskoeffizienten im Falle einer einzigen Ursache. Bei Ausschaltung von u ist der Zusammenhang zwischen a und v der maximal starke, und bei Ausschaltung von v ist der Zusammenhang zwischen a und u der maximal starke.

Falls ausser u noch eine Nicht-Ursache w ausgeschaltet wird, folgt aus (23) und (27), dass auch dann wieder der Zusammenhang zwischen a und v der maximal-starke ist. Falls ausser v noch eine Nicht-Ursache w ausgeschaltet wird, folgt ebenso aus (23) und (27), dass dann der Zusammenhang zwischen a und u der maximal-starke ist.

γ . *Fall der Wirksamkeit von n Ursachen.* Auf Grund von Formel (23) für partielle Korrelationskoeffizienten dritter Ordnung sowie auf Grund von Formel (24) für partielle Korrelationskoeffizienten vierter Ordnung und auf Grund entsprechender Formeln für Korrelationskoeffizienten von noch höherer als vierter Ordnung ergibt sich in derselben Weise ganz allgemein, dass bei Wirksamkeit von n Ursachen

- 1) die Ausschaltung von $n - 1$ dieser Ursachen zu partiellen Korrelationskoeffizienten $(n - 1)$ -ter Ordnung führt, bei denen die n -te Ursache (d.h. die nicht ausgeschaltete Ursache) den maximal-starken Zusammenhang aufweist,
- 2) die Ausschaltung von $n - 1$ dieser Ursachen und irgendeiner Nicht-Ursache zu partiellen Korrelationskoeffizienten n -ter Ordnung führt, bei denen auch wieder die n -te Ursache (die nicht ausgeschaltete Ursache) den maximal-starken Zusammenhang aufweist. — Entsprechendes gilt, wenn neben den $n - 1$ Ursachen nicht nur *eine* Nicht-Ursache, sondern mehrere Nicht-Ursachen ausgeschaltet werden.

Im Sonderfall $n = 1$ hat man den maximal-starken Zusammenhang, der im allgemeinen erst nach Punkt 1) gesucht werden muss, bereits durch das Maximum in der Reihe der gewöhnlichen Korrelationskoeffizienten gegeben.

8. *Über die Anwendbarkeit des Kriteriums.* Das Kriterium des maximal-starken Zusammenhangs ist gewiss unter der Voraussetzung anwendbar, dass im jeweiligen statistischen Untersuchungsmaterial alle n wirksamen Ursachen enthalten sind (d.h. nicht nur indirekt wirken). Es fragt sich, wie das Kriterium in einem Fall arbeiten würde, in dem nicht alle n wirksamen Ursachen im Material enthalten sind.

Beim Verfahren der nachträglichen Ursachen-Bestimmung, bei dem das Kriterium verschwindender Korrelationskoeffizienten benutzt wird, wäre es von vornherein klar, wie die Dinge in einem solchen Fall liegen: Falls nicht alle n Ursachen im Material enthalten sind, gibt es keinen verschwindenden partiellen Korrelationskoeffizienten.

Beim Verfahren der direkten Ursachen-Bestimmung bietet sich mit Hilfe des Kriteriums des maximal-starken Zusammenhangs eine charakteristische Ringbildungs-Situation dar, vorausgesetzt dass alle n wirksamen Ursachen im Material enthalten sind. Fehlt eine Ursache, sagen wir die n -te Ursache, so kann durch Ausschaltung der übrigen $n - 1$ Ursachen keine charakteristische Ringbildungs-Situation entstehen, weil die n -te Ursache stets indirekt über die Nicht-Ursachen wirkt, aber über keine von ihnen vollständig. Ausschaltung einer Nicht-Ursache ersetzt also nicht die Ausschaltung der n -ten Ursache, und andererseits gibt es keine Nicht-Ursache, deren Ausschaltung ohne Einfluss auf die Entstehung charakteristischer Situationen wäre. Nach Punkt 2) des Kriteriums muss aber eine charakteristische Situation, die nach Punkt 1) des Kriteriums zustandekommt, auch bei zusätzlicher Ausschaltung von Nicht-Ursachen auftreten, d.h. die zusätzliche Ausschaltung von Nicht-Ursachen muss ohne Einfluss auf das Auftreten der charakteristischen Situation bleiben.

II. Zur statistischen Sicherheit des Kriteriums des maximal-starken Zusammenhangs.

α. Punkt 1) des Kriteriums. Nach Punkt 1) ergeben sich charakteristische Sachlagen, sobald n mindestens gleich 2 ist. Bei 2 Ursachen u und v ergibt die Ausschaltung der Ursache u , dass der maximal-starke Zusammenhang in der Reihe der entstehenden partiellen Korrelationskoeffizienten gerade der Zusammenhang zwischen v und dem Untersuchungs-Objekt ist, während Ausschaltung von v ergibt, dass der maximal-starke Zusammenhang in der Reihe der entstehenden partiellen Korrelationskoeffizienten gerade der Zusammenhang zwischen u und dem Untersuchungs-Objekt ist. Bei 3 Ursachen u , v , w ergibt die Ausschaltung von u und v , dass der maximal-starke Zusammenhang in der Reihe der entstehenden partiellen Korrelationskoeffizienten zweiter Ordnung gerade der Zusammenhang zwischen w und dem

Untersuchungs-Objekt ist ; ebenso ergibt die Ausschaltung von u und w , dass der maximal-starke Zusammenhang in der Reihe der entstehenden partiellen Korrelationskoeffizienten zweiter Ordnung gerade der Zusammenhang zwischen v und dem Untersuchungs-Gegenstand ist ; und ebenso ergibt die Ausschaltung von v und w , dass der maximal-starke Zusammenhang in der Reihe der entstehenden partiellen Korrelationskoeffizienten zweiter Ordnung gerade der Zusammenhang zwischen u und dem Untersuchungs-Gegenstand ist. Bei mehr als 3 Ursachen gilt entsprechendes.

Man kann sich leicht überlegen, wie gross die Wahrscheinlichkeit dafür ist, dass die besprochenen Sachlagen rein zufällig entstehen könnten. Es sei m die Anzahl der Beobachtungs-Objekte, die neben dem Untersuchungs-Gegenstand herangezogen werden. Wenn sich nun bei Ausschaltung eines Objekts u ergibt, dass ein Objekt v den maximal-starken Zusammenhang in der Reihe der partiellen Korrelationskoeffizienten aufweist, so besteht bei Ausschaltung von v die Wahrscheinlichkeit $\frac{1}{m-1}$ dafür, dass u den maximal-starken Zusammenhang in der Reihe der partiellen Korrelationskoeffizienten aufweist. Hiermit besteht also die Wahrscheinlichkeit $\frac{1}{m-1}$ für zufälliges Entstehen einer charakteristischen Sachlage mit u und v .

Wenn sich bei Ausschaltung von u und v ein Objekt w als Objekt mit maximal-starkem Zusammenhang ergibt, so besteht bei Ausschaltung von u und w die Wahrscheinlichkeit $\frac{1}{m-2}$ dafür, dass v als Objekt mit maximal-starkem Zusammenhang auftritt, und es besteht bei Ausschaltung von v und w auch die Wahrscheinlichkeit $\frac{1}{m-2}$ dafür, dass u als Objekt mit maximal-starkem Zusammenhang auftritt. Es besteht also die Wahrscheinlichkeit $\frac{1}{(m-2)^2}$ für zufälliges Entstehen einer charakteristischen Sachlage mit u, v, w . Allgemein besteht die Wahrscheinlichkeit $\frac{1}{(m-n+1)^{n-1}}$ für zufälliges Entstehen einer charakteristischen Sachlage mit n Objekten.

Wenn, wie im praktischen Beispiel der Arbeit, $m = 21$ und $n = 3$ ist, wird $\frac{1}{(m - n + 1)^{n-1}} = \frac{1}{19^2} = 0,00277$. Das ist etwa die Vertrauensgrenze, die S. KOLLER in seinen bekannten graphischen Tafeln annimmt. Englische Statistiker benutzen als Vertrauensgrenze den Wert 0,01 oder sogar auch 0,05.

β. *Punkt 2) des Kriteriums.* Die Heranziehung von Punkt 2) des Kriteriums ermöglicht eine Prüfung, ob das mit Hilfe von Punkt 1) gewonnene Ergebnis so bedeutungsvoll ist, dass es bestätigt werden kann.

Im Falle einer *einzigsten Ursache* u besteht die Wahrscheinlichkeit $\frac{1}{m}$ dafür, dass u den maximal-starken Zusammenhang in der Reihe der gewöhnlichen Korrelationskoeffizienten zufällig aufweist. Bei Ausschaltung eines Objekts, das nicht u ist, besteht die Wahrscheinlichkeit $\frac{1}{m-1}$ dafür, dass u den maximal-starken Zusammenhang in der Reihe der zustandekommenden partiellen Korrelationskoeffizienten zufällig aufweist.

Für die Ausschaltung eines Objekts, das nicht u ist, können verschiedene Objekte herangezogen werden. In jedem Falle besteht die Wahrscheinlichkeit $\frac{1}{m-1}$ dafür, dass u den maximal-starken Zusammenhang in der Reihe der zustandekommenden partiellen Korrelationskoeffizienten erster Ordnung aufweist.

Im Falle von *2 oder mehreren Ursachen* bekommt man eine entsprechende Wahrscheinlichkeit dafür, dass u zufällig den maximal-starken Zusammenhang aufweist.

Nach den Ausführungen des vorigen Abschnitts besteht z.B. für 2 Ursachen u und v die Wahrscheinlichkeit $\frac{1}{m-1}$ dafür, dass Ausschaltung von u zu einem maximal-starken Zusammenhang von v führt. Die Ausschaltung von u zusammen mit einer Nicht-Ursache und das Auftreten von v im maximal-starken Zusammenhang der entstehenden Reihe partieller Korrelationskoeffizienten zweiter Ordnung hat bei zufälligem Zustandekommen die Wahrscheinlichkeit $\frac{1}{m-2}$. Die gleiche Wahrschein-

lichkeit besteht dafür, dass die Ausschaltung von v zusammen mit einer Nicht-Ursache zu partiellen Korrelationskoeffizienten zweiter Ordnung führt, bei denen der maximal-starke Zusammenhang der Zusammenhang mit u ist. Die Gesamt-Wahrscheinlichkeit dafür, dass die Ausschaltung von u zusammen mit einer Nicht-Ursache zu einem maximal-starken Zusammenhang mit v und die Ausschaltung von v zusammen mit einer Nicht-Ursache zu einem maximal-starken Zusammenhang mit u führt, beträgt bei reiner Zufallsmässigkeit $\frac{1}{(m-2)^2}$. Allgemein besteht die entsprechende

Wahrscheinlichkeit $\frac{1}{(m-n)^n}$ für beliebiges n . Wenn z.B. $m = 21$

und $n = 3$ ist, wird $\frac{1}{(m-n)^n} = \frac{1}{18^3} = 0,00017$.

Die Gesamt-Wahrscheinlichkeit für eine zufällige charakteristische Sachlage nach Punkt 1) und Punkt 2) des Kriteriums ist das Produkt der Wahrscheinlichkeit nach Punkt 1) und der Wahrscheinlichkeit nach Punkt 2).

III. Praktisches Beispiel. Das praktische Beispiel, das in Teil 1 der Arbeit und im Haupt-Abschnitt « Verfahren der nachträglichen Ursachen-Bestimmung » herangezogen wurde, möge hier weiterbehandelt werden. Zur Anwendung des Kriteriums des maximal-starken Zusammenhangs ist es nötig, aus der Reihe der berechneten Korrelationskoeffizienten (gewöhnliche sowie partielle Korrelationskoeffizienten verschiedener Ordnung) den jeweils maximal-starken Zusammenhang herauszusuchen und das gefundene Objekt des maximalen Zusammenhangs stets dem oder den ausgeschalteten Objekten gegenüberzustellen. Tabelle 7 zeigt die ersten 7 Werte jeder berechneten Reihe. Die zur Ausschaltung kommenden Objekte sind die gleichen wie in Teil 1 der Arbeit und wie im Haupt-Abschnitt « Verfahren der nachträglichen Ursachen-Bestimmung », d.h. es wird für die Reihenfolge der Ausschaltung nicht die Reihe der gewöhnlichen Korrelationskoeffizienten zugrundegelegt, sondern die entsprechende Reihe der durchschnittlichen partiellen Korrelationskoeffizienten erster Ordnung, die am rechten Rand von Tabelle 2 aufgeführt ist.

Tabelle 8 stellt das jeweilige Objekt mit maximal-starkem Zusammenhang in besonders übersichtlicher Form den ausge-

TABELLE 7: Die Korrelationskoeffizient zwischen der Krebssterblichkeit und dem jeweils aufgeführten Objekt, zunächst ohne Ausschaltung, dann für verschiedene Ausschaltungen. Die ausgeschalteten Objekte sind in der ersten Spalte links angegeben.

ausgeschaltet:

nichts	tF 0,890	tP 0,884	Energie 0,846	kW 0,843	Telefon 0,829	Kreislauf Sterblk. 0,825	A 0,815
tF.	kW 0,500	A 0,461	Energie 0,381	Zeitung 0,332	Kreislauf Sterblk. 0,320	Radio 0,317	Zement 0,309
tP.	A 0,578	veget. 0,392	kW 0,369	tF 0,356	Kreislauf Sterblk. 0,329	Energie 0,324	Zement 0,260
A	tP 0,752	tF 0,716	Telefon 0,704	kW 0,688	Radio 0,675	Energie 0,670	raff. Zucker 0,641
kW	tF 0,679	A 0,623	tP 0,589	Kreislauf Sterblk. 0,487	t. Kohle- hydr. 0,474	veget. 0,440	raff. Zucker 0,360
tF, tP. . . .	A 0,506	kW 0,430	veget. 0,325	Energie 0,301	Kreislauf Sterblk. 0,265	Zement 0,247	Zeitung 0,228
tF, A	kW 0,480	Radio 0,413	Zeitung 0,378	tP 0,366	Energie 0,359	Kreislauf Sterblk. 0,268	Ärzte 0,252
tF, kW . . .	A 0,438	veget. 0,316	Kreislauf Sterblk. 0,166	Zeitung 0,126	Säugling. Überleb. 0,122	Radio 0,102	Export 0,087
tP, A	kW 0,321	Radio 0,271	Energie 0,240	Zeitung 0,207	Telefon 0,183	t. Kohle- hydr. 0,166	raff. Zucker 0,166
tP, kW . . .	A 0,556	tF 0,420	veget. 0,403	Kreislauf Sterblk. 0,262	Zement 0,108	t. Kohle- hydr. 0,108	Briefe 0,092
A , kW	tF 0,536	tP 0,511	t. Kohle- hydr. 0,490	raff. Zucker 0,408	Radio 0,384	Kreislauf Sterblk. 0,342	Telefon 0,326
tF, tP, A . .	kW 0,360	Radio 0,288	Energie 0,243	Zeitung 0,242	Kreislauf Sterblk. 0,190	Telefon 0,172	raff. Zucker 0,138
tF, tP, kW .	A 0,454	veget. 0,324	Kreislauf Sterblk. 0,164	Zeitung 0,123	Säugling. Überleb. 0,118	Radio 0,097	Export 0,081
tF, A , kW .	Radio 0,224	t. Kohle- hydr. 0,194	Zeitung 0,191	tP 0,137	raff. Zucker 0,124	Kreislauf Sterblk. 0,121	Ärzte 0,082
tP, A , kW .	tF 0,232	t. Kohle- hydr. 0,217	Radio 0,183	raff. Zucker 0,148	Kreislauf Sterblk. 0,147	Zeitung 0,120	Ärzte 0,032

TABELLE 8: Gegenüberstellung der jeweils ausgeschalteten Objekte und des sich ergebenden Objekts mit maximal-starkem Zusammenhang. Das letztere ist durch Unklammerung mit runden Klammern gekennzeichnet.

	tF	tP	A	kW	
tF	tF	(tF)	(tP)	tP	A
tP	A	(kW)	(A)	kW	kW
	tF	tF	(tF)	tP	
	tP	tP	(A)	(A)	
	(A)	kW	(kW)	(kW)	
		(tF)			
		tP			
		(A)			
		(kW)			

geschalteten Objekten gegenüber. In der ersten Zeile befinden sich die 4 ersten Objekte der Ausschaltungs-Anordnung. In der zweiten Zeile befinden sich alle denkbaren Kombinationen von je zweien dieser Objekte, in der dritten Zeile alle denkbaren Kombinationen von je dreien dieser Elemente und in der vierten Zeile die einzige Kombination aller 4 Elemente. Die Gegenüberstellung der jeweils ausgeschalteten Objekte und des jeweils erscheinenden Objekts mit maximal-starkem Zusammenhang geschieht folgendermassen: Nach Tabelle 7 ergibt sich bei Ausschaltung von tF das Objekt kW als Objekt mit maximal-starkem Zusammenhang; infolgedessen ist in Tabelle 8 das Objekt kW in der Kombination $\begin{smallmatrix} tF \\ kW \end{smallmatrix}$ durch Umklammerung mit runden Klammern gekennzeichnet. Nach Tabelle 7 ergibt Ausschaltung von kW das Objekt tF als Objekt mit maximal-starkem Zusammenhang; infolgedessen ist in Tabelle 8 das Objekt tF in der Kombination $\begin{smallmatrix} tF \\ kW \end{smallmatrix}$ durch Umklammerung gekennzeichnet. Nach Tabelle 7 hat bei Ausschaltung von tP das Objekt A den maximal-starken Zusammenhang; also ist in Tabelle 8 das Objekt A in der Kombination $\begin{smallmatrix} tP \\ A \end{smallmatrix}$ durch Umklammerung gekennzeichnet. Nach Tabelle 7 hat bei Ausschaltung von A das Objekt tP den maximal-starken Zusammenhang; also ist in Tabelle 8 in der

Kombination $\begin{smallmatrix} tP \\ A \end{smallmatrix}$ das Objekt tP durch Umklammerung gekennzeichnet. Nach Tabelle 7 hat bei Ausschaltung von tF und tP das Objekt A den maximal-starken Zusammenhang; also ist in Tabelle 8 in der Kombination $\begin{smallmatrix} tF \\ tP \\ A \end{smallmatrix}$ das Objekt A durch Umklammerung gekennzeichnet. Die sonstige in Tabelle 8 vorkommende Kennzeichnung, durch Umklammerung ist auf gleiche Weise entstanden.

Punkt 1 des Kriteriums des maximalen Zusammenhangs ist dann erfüllt, wenn es in einer Zeile genau eine Kombination gibt, deren Objekte sämtlich umklammert sind. Dies ist offensichtlich in der dritten Zeile der Fall; dort ist $\begin{smallmatrix} tF \\ A \\ k W \end{smallmatrix}$ die einzige Kom-

bination, in welcher sämtliche Objekte umklammert sind. Hiernach wären tF , A und $k W$ die gesuchten Ursachen. Unter der Hypothese, dass dies der Fall ist, also tP keine Ursache ist, zeigt sich, dass dann auch Punkt 2 des Kriteriums erfüllt ist. Aus der letzten Zeile der Tabelle 8 ergibt sich nämlich: Werden tF , tP und A ausgeschaltet, so zeigt $k W$ den maximal-starken Zusammenhang, ebenso wie wenn dieselbe Ausschaltung ohne tP erfolgt. Werden tF , tP und $k W$ ausgeschaltet, so zeigt A den maximal-starken Zusammenhang, ebenso wie wenn dieselbe Ausschaltung ohne tP erfolgt. Werden tP , A und $k W$ ausgeschaltet, so zeigt tF den maximal-starken Zusammenhang, ebenso wie wenn dieselbe Ausschaltung ohne tP erfolgt.

Ein rein zufallsmässiges Zustandekommen des Ergebnisses hätte im Hinblick auf Punkt 1) und Punkt 2) des Kriteriums die Wahrscheinlichkeit $\frac{1}{(m-n+1)^{n-1}} \cdot \frac{1}{(m-n)^n}$, die bei Einsetzung von $m = 21$ und $n = 3$ den sehr kleinen Wert 0,00000047 aufweist.

Das Resultat, das hier mit Hilfe des Kriteriums des maximalen Zusammenhangs gewonnen wurde, ist also genau dasselbe wie das maximal-wahrscheinliche Ergebnis, das sich bei Benutzung des Kriteriums verschwindender Korrelationskoeffizienten herausstellt. Diese Übereinstimmung ist insofern beachtenswert, weil hierbei mit ganz verschieden gearteten Grössen gearbeitet

wird. Während in dem einen Falle mit dem Verschwinden des Durchschnitts zahlreicher Korrelationskoeffizienten gearbeitet wird, stützt sich das andere Verfahren auf das Maximum solcher Korrelationskoeffizienten.

IV. Fingiertes Beispiel. Das fingierte Beispiel, das im Haupt-Abschnitt « Verfahren der nachträglichen Ursachen-Bestimmung » behandelt wurde, möge das Gesamtbild durch Behandlung mit dem Kriterium des maximalen Zusammenhangs vervollständigen. Tabelle 9 gibt die Korrelationskoeffizienten zwi-

TABELLE 9: Die Korrelationskoeffizienten zwischen *a* und den jeweils aufgeführten Objekten, zunächst ohne Ausschaltung, dann für verschiedene Ausschaltungen

ausgeschaltet :

nichts	b	u	c	v	d	e
	0,715	0,700	0,650	0,600	0,585	0,520
u	v	b	c	d	e	
	0,840	0,553	0,485	0,423	0,367	
b	u	c	d	v	e	
	0,525	0,540	0,380	0,354	0,327	
v	u	b	c	d	e	
	0,875	0,576	0,505	0,441	0,382	
u, b	v	c	d	e		
	0,759	0,357	0,305	0,260		
u, v						
	0,000	0,000	0,000	0,000		
b, v	u	c	d	e		
	0,806	0,378	0,324	0,276		

schen *a* und dem jeweils aufgeführten Objekt wieder, zunächst ohne Ausschaltung, dann bei Ausschaltung von *u*, bei Ausschaltung von *b*, bei Ausschaltung von *v*, ferner von *u b*, von *u v* und von *b v*.

Tabelle 10 stellt die jeweils ausgeschalteten Objekte dem sich ergebenden Objekt mit maximal-starkem Zusammenhang gegenüber. Das Objekt mit maximal-starkem Zusammenhang wird in der gleichen Weise durch Umklammerung gekennzeichnet wie in Tabelle 8. Wie man sieht, ist Punkt 1 des Kriteriums

TABELLE 10: Gegenüberstellung der jeweils ausgeschalteten Objekte und des sich ergebenden Objekts mit maximal-starkem Zusammenhang

u	b	v
(u)	(u)	b
b	(v)	v
(u)		
b		
(v)		

des maximalen Zusammenhangs für die Kombination $\begin{smallmatrix} u \\ v \end{smallmatrix}$ in der zweiten Zeile erfüllt. Unter der Hypothese, dass u und v die vorkommenden Ursachen sind, ist Punkt 2 des Kriteriums gleichfalls erfüllt, wie Zeile 3 vor Augen führt.

Das hier erhaltene Resultat ist genau dasselbe wie auf Grund des Verfahrens der nachträglichen Ursachenbestimmung und befindet sich auch mit der Konstruktion des fingierten Beispiels in vollstem Einklang.

V. Erweiterung des Wirkungsbereichs des Kriteriums. Während das Verfahren, das den verschwindenden Korrelationskoeffizienten benutzt, mit seinem Schwergewicht im Anwendungsgebiet der Wissenschaft vom Zufall liegt, befindet sich das Verfahren, das den maximal-starken Zusammenhang verwendet, mit seinem Schwergewicht mehr in einem Gebiet, in dem das Auftreten systematischer Gegebenheiten die Hauptrolle spielt und demgegenüber das Spiel des Zufalls weniger in die Waagschale fällt. Beim Verfahren mit dem maximal-starken Zusammenhang werden die gesuchten Ursachen auf dem Wege über den maximal-starken Korrelationskoeffizienten ausfindig gemacht, d.h. nach mathematisch-statistischen Begriffen wäre hier die Sicherheit der Differenz zwischen dem maximal-starken und dem zweitstärksten Korrelationskoeffizienten zu beurteilen. Aber einerseits kommt es bei Vorhandensein mehrerer Ursachen gar nicht darauf an, in welcher Reihenfolge sie ausfindig gemacht werden, und wenn sich eine Ursache korrelationsmässig nur geringfügig von einer Nicht-Ursache unterscheidet, ist es auch kein grosses Unglück, falls die Unterscheidung nicht gelingt. Ande-

rerseits sind zwei hohe Korrelationskoeffizienten gewöhnlich auch so stark miteinander korreliert, dass der Zufallsfehler ihrer Differenz ziemlich klein wird.

Es möge nun noch eine Frage in den Vordergrund gerückt werden, die für die praktischen Belange wichtig ist. Wenn das Kriterium des maximal-starken Zusammenhangs eine charakteristische Sachlage ausfindig macht, so besteht schon bei Heranziehung von mässig vielen Beobachtungsobjekten nur noch eine praktisch bedeutungslose Wahrscheinlichkeit dafür, dass das Ergebnis rein zufallsmässig entstanden sein könnte. Man braucht also im allgemeinen keine Sorge darum zu haben, dass man mehr findet, als man finden will. Eher könnte es sein, dass man zu wenig ausfindig macht. Um einem solchen möglichen Übel abzuweichen, kann man das Kriterium des maximal-starken Zusammenhangs in zweckmässiger Weise erweitern.

Die Erweiterung besteht darin, dass man nicht nur die maximal-starken Zusammenhänge, sondern auch die zweitstärksten Zusammenhänge in die Betrachtung einbezieht. Die Wahrscheinlichkeit, dass eine charakteristische Sachlage dann rein zufallsmässig entsteht, ist natürlich grösser als bei alleiniger Betrachtung der maximal-starken Zusammenhänge, und zwar doppelt so gross im Falle einmaliger Erweiterung, 2^2 mal so gross im Falle zweimaliger Erweiterung, usw.

Falls man mit Hilfe des erweiterten Verfahrens noch eine zweite charakteristische Sachlage ausfindig macht, nachdem schon mit Hilfe des einfachen Verfahrens eine charakteristische Sachlage gefunden worden ist, so fällt die Entscheidung durch Punkt 2 des Kriteriums und eventuell unter Berücksichtigung von Wahrscheinlichkeiten über zufallsmässiges Zustandekommen; im letzten Fall können wir mit Hilfe des Bayesschen Theorems die Wahrscheinlichkeiten dafür berechnen, mit denen die Hypothesen für verschiedene charakteristische Sachlagen Gültigkeit haben.

Die Behandlung des praktischen Beispiels zeigt, dass das Verfahren mit dem maximal-starken Korrelationskoeffizienten eine schärfere Hypothesen-Trennung ermöglicht als das Verfahren mit dem verschwindenden Korrelationskoeffizienten. Während beim Verfahren mit dem verschwindenden durchschnittlichen Korrelationskoeffizienten die Werte von 9 Hypothesen eine Vereinbarkeit mit dem Wert Null aufweisen (weil der jeweils zu-

gehörige Quotient gemäss Tabelle 3 zwischen +3 und -3 liegt), ist es beim Verfahren mit dem maximal-starken Korrelationskoeffizienten eine einzige Hypothese, die mit dem praktischen Befund vereinbar ist.

Zur Veranschaulichung des Gesagten möge das oben behandelte *praktische Beispiel* dienen. Man entnimmt die Objekte mit zweitstärksten Zusammenhängen, ebenso wie die mit maximal-starken Zusammenhängen, wieder aus Tabelle 7. Eine Gegenüberstellung der jeweils ausgeschalteten Objekte und des sich ergebenden Objekts mit zweitstärkstem Zusammenhang wie auch des schon in Tabelle 8 aufgeführten Objekts mit maximal-starkem Zusammenhang gibt Tabelle 10.

TABELLE 11: Gegenüberstellung der jeweils ausgeschalteten Objekte und des sich ergebenden Objekts mit maximal-starkem Zusammenhang wie auch des Objekts mit zweitstärkstem Zusammenhang. Das letztere ist durch Umlammerung mit eckigen Klammern gekennzeichnet.

	tF	tP	A	kW	
tF	$[tF]$	(tF)	(tP)	tP	$[A]$
tP	$[A]$	(kW)	(A)	kW	kW
	tF	$[tF]$	(tF)	$[tP]$	
	tP	tP	(A)	(A)	
	(A)	$[kW]$	(kW)	(kW)	
		(tF)			
		tP			
		(A)			
		(kW)			

Dabei sind die Objekte mit zweitstärkstem Zusammenhang durch Umlammerung mit eckigen Klammern gekennzeichnet, während die Kennzeichnung der Objekte mit maximal-starkem Zusammenhang wie in Tabelle 8 durch runde Klammern erfolgt.

Wie man sieht, ergibt sich zusätzlich die charakteristische

$[tP]$

Sachlage (A) mit einer eckigen Klammer und ausserdem die (kW)

charakteristische Sachlage $[tF]$
 $[A]$ mit zwei eckigen Klammern.

Keine von beiden wird jedoch nach Punkt 2) des Kriteriums
 $[tP]$
 verifiziert. Zur Verifizierung von (A) wäre es nötig, dass die
 (kW)

Objekte tP , A und kW bei Ausschaltungen höherer Ordnung stets mit runder Umklammerung auftreten, was, wie man an Hand des Quadrupels der vierten Zeile erkennt, nicht der Fall ist. Dies bedeutet zwar, dass zwischen tP und der Krebssterblichkeit kein direkter Zusammenhang besteht, doch spricht der Befund dafür, dass tP immerhin recht eng mit tF , kW und A zusammenwirkt. — Zur Verifizierung von $[tF]$
 $[A]$ wäre es nötig, dass die Objekte tF und A bei Ausschaltungen höherer Ordnung stets mit runder Umklammerung auftreten, was, wie die erste Kombination der dritten Zeile zeigt, nicht der Fall ist.

Erwähnt sei daneben, dass die charakteristische Sachlage (kW)
 (tF)
 infolge der Berücksichtigung eckiger Klammern bedeutungsvoller geworden ist. Diese Sachlage tritt zwar bei Ausschaltungen höherer Ordnung nicht durchweg mit runden Klammern auf, sie tritt aber durchweg umklammert auf, wenn man runde und eckige Klammern berücksichtigt. Das spricht für erhebliche Bedeutung der Objekte tF und kW , und es stimmt auch mit einem Ergebnis auf Grund des Verfahrens verschwindender Korrelationskoeffizienten überein: Nach Tabelle 3 besitzt die Ausschaltung kW
 tF mit dem fast verschwindenden Quotienten $+0,03$ die höchste Wahrscheinlichkeit unter den Ausschaltungen zweier Objekte.

Das *fingierte* Beispiel, dessen Zahlen in Tabelle 9 niedergelegt sind, ergibt bei Erweiterung der Methodik auf die zweitstärksten Zusammenhänge ein Bild, das durch Tabelle 11 vor Augen geführt wird. Wie man sieht, ergibt sich zusätzlich die charakteristische Sachlage (u)
 $[b]$. Zur Verifizierung nach Punkt 2) des Kriteriums müssten die Objekte u und b bei Ausschaltungen höherer Ordnung stets mit Umklammerung (wenigstens bei gemeinsamer Berücksichtigung runder und eckiger Klammern) auftreten. Das ist jedoch nicht der Fall, wie man an Hand des in der dritten Zeile
 (u)
 stehenden Tripels b erkennt.
 (v)

TABELLE 12 : Gegenüberstellung der jeweils ausgeschalteten Objekte und des sich ergebenden Objekts mit maximal-starkem Zusammengang wie auch des Objekts mit zweitstärkstem Zusammenhang. Das letztere ist durch eckige Klammern gekennzeichnet.

u	b	v
(u)	(u)	[b]
[b]	(v)	v

(u)
b
(v)

Zusammenfassung

Vorausgesetzt wird eine Menge von Beobachtungs-Objekten (z.B. das Durchschnitts-Alter einer Bevölkerung, die durchschnittliche Einkommens-Höhe, die Zahl der Personen-Kraftwagen pro Kopf, usw.), die mehr oder weniger starke statistische Zusammenhänge mit einem Untersuchungs-Objekt (z.B. der Krebssterblichkeit) aufweisen. Unter Zugrundelegung von partiellen Korrelationskoeffizienten verschiedenster Ordnung wird eine Methode entwickelt, mit der erstens geprüft werden kann, ob sich in der Menge der betrachteten Beobachtungs-Objekte alle praktisch belangreichen Objekte befinden, die einen direkten Zusammenhang (d.h. einen Kausal-Zusammenhang) mit dem Untersuchungs-Objekt besitzen, und mit deren Hilfe zweitens erforscht werden kann, welches genau die praktisch belangreichen Objekte sind, die einen direkten Zusammenhang mit dem Untersuchungs-Objekt haben.

Als Kriterium für die erstgenannte Prüfung dient das Null-Werden eines arithmetischen Mittels partieller Korrelationskoeffizienten, wobei die Beurteilung der Sicherheit des Resultats durch empirische und theoretische mathematisch-statistische Hilfsmittel erfolgt. Auf gleichem Wege kann auch eine Erkundung der einzelnen direkten Zusammenhänge erfolgen ; die dabei erzielbare Hypothesen-Trenn-Schärfe ist jedoch relativ schlecht.

Eine relativ gute Hypothesen-Trenn-Schärfe ist auf einem andern Weg möglich, auf welchem systematische Zuordnungen der Beobachtungs-Objekte zu maximal-starken partiellen Korrelationskoeffizienten benutzt werden und das Auftreten charakteristischer Ringbildungs-Situationen erkennen lässt, welche Zu-

sammenhänge die direkten sind. Die Beurteilung der Sicherheit erfolgt hierbei mit Hilfsmitteln der Wahrscheinlichkeitsrechnung.

Die Methodik wird an Hand eines praktischen Beispiels und ausserdem an Hand eines fingierten Beispiels demonstriert.

Als praktisches Beispiel wird die Krebssterblichkeit des Menschen unter Heranziehung der Statistik der United Nations untersucht. Es ergeben sich 3 direkte Zusammenhänge, nämlich in erster Linie ein Zusammenhang mit den tierischen Fett-Kalorien der Ernährung pro Kopf der Bevölkerung, zweitens mit dem Kilowattstunden-Verbrauch pro Kopf und drittens mit dem Durchschnitts-Alter der Bevölkerung. Dabei ist der Zusammenhang zwischen Krebssterblichkeit und tierischen Fett-Kalorien plausibel, weil die tierischen Fette durchschnittlich einen höheren Sättigungs-Grad als die vegetarischen Fette haben; ein höherer Sättigungs-Grad bedeutet verringerte Zell-Atmung und damit erhöhte Krebs-Wahrscheinlichkeit. Der Zusammenhang zwischen Krebssterblichkeit und Kilowattstunden-Verbrauch erscheint zunächst etwas verwunderlich. Aber wenn man sich vergegenwärtigt, dass Wärme-Grade über 42° Celsius eine Erhöhung des Sättigungs-Grades von Fetten bewirken, scheint die Wärme-Wirkung des Koch-Stromes offenbar eine Komponente des Kilowattstunden-Verbrauchs zu sein, die einen direkten Zusammenhang mit der Krebssterblichkeit hat. Möglicherweise spielen auch andere Komponenten des Kilowattstunden-Verbrauchs eine Rolle, so dass als Krebs-Ursache eine Störung des bio-elektrischen Zustandes der Körper-Zellen anzusprechen wäre. Mit diesem Gesichtspunkt würde die übliche Meinung harmonisieren, dass bei der Zell-Atmung eine Elektronen-Wanderung stattfindet und dass schnell wachsendes Gewebe ein auffällig hohes Potential aufweist. — Aus dem Resultat folgt, dass die 3 gefundenen Zusammenhänge die einzigen *praktisch* ins Gewicht fallenden direkten Zusammenhänge mit der Krebssterblichkeit sind. Eine Existenz anderer, praktisch unwichtiger direkter Zusammenhänge ist dabei nicht ausgeschlossen. Falls andere Zusammenhänge praktisch ins Gewicht fallen, sind es demnach indirekte Zusammenhänge, wozu auch Zusammenhänge zwischen der Krebssterblichkeit und Objekten gehören, die eine effektsteigernde Wirkung auf die Entwicklung der Ursachen ausüben.

Das fingierte Beispiel führt vor Augen, dass die Methodik auch bei Vorliegen anormaler Situationen gut funktioniert.

S u m m a r y

A statistical method of finding causal relationships is developed by using partial correlation coefficients. Regarding an object of research and a quantity of other objects which are statistically associated with each other as well as with the object of research, we can find out with the aid of the method whether there are objects that have direct relationships with the object of research. Farther we can find out which of them are these very objects. In order to increase the significance of the result still another way is described allowing to get the same result by other means.

Applications of the twofold method are demonstrated by a practical example (cancer in man) and by an artificial example.

INHALT

	Seite
Einleitung	3
TEIL I.	4
A. <i>Monokausal-Theorie</i>	4
I. Eine Ursache und 2 abhängige Variable	4
II. Eine Ursache und 3 abhängige Variable	5
B. <i>Polykausal-Theorie</i>	6
I. Frage der Identität zwischen maximal-starkem und direktem Zusammenhang	6
II. Hinreichende Bedingungen für Identität zwischen maximal-starkem und direktem Zusammenhang	7
III. Das Kriterium verschwindender partieller Korrelationskoeffizienten	10
C. <i>Ein praktisches Beispiel</i>	11
I. Das statistische Material	11
II. Die berechneten Korrelationskoeffizienten.	11
III. Die statistische Sicherheit des Ergebnisses	16
TEIL 2.	18
D. <i>Verfahren der nachträglichen Ursachen-Bestimmung</i>	18
I. Die verschiedenen Möglichkeiten der Ausschaltung von α aus n Grössen	19
II. Zur Durchführung der erforderlichen Rechnungen	20
III. Praktisches Beispiel	20
IV. Ein fingiertes Beispiel	22
E. <i>Verfahren der direkten Ursachen-Bestimmung</i>	26
I. Das Kriterium des maximal-starken Zusammenhangs	26
II. Zur statistischen Sicherheit des Kriteriums des maximal-starken Zusammengangs	29
III. Praktisches Beispiel	32
IV. Fingiertes Beispiel	36
V. Erweiterung des Wirkungsbereichs des Kriteriums.	37
Zusammenfassung	41
Summary.	43

CAMILO DAGUM

Teoría de la transvariación Sus aplicaciones a la Economía

I. INTRODUCCION

1.1 — El desarrollo del Método Estadístico y de su fundamento matemático, la Teoría de la Probabilidad, de los que nos ocupamos precedentemente, nos provee un eficazísimo aparato metodológico para el tratamiento cuantitativo de los colectivos y de sus respectivas muestras de observaciones, sean que las mismas formen una serie del tiempo o una serie de frecuencia y, para estas últimas, tanto para el caso de referirse a atributos cuantitativos como para el de atributos cualitativos.

Las series del tiempo recibieron una adecuada y rigurosa sistematización, que en el campo económico se traduce en la consideración analítica de sus cuatro grandes componentes que concurren a la determinación de los valores observados a través del tiempo y que son: la componente tendencial, la componente cíclica, la componente estacional y la componente aleatoria.

En el dominio de las series de frecuencia encontramos también una rigurosa y fecunda sistematización que se inicia con las distribuciones clásicas de frecuencias con que se las puede representar, siguiendo con las fórmulas provistas para la determinación del reducido número de parámetros que las sintetizan y caracterizan, como las medidas del valor medio, de la variabilidad, asimetría y kurtosis y concluyendo, tanto para las series de frecuencia como para las series del tiempo, con el notable instrumento metodológico que conocemos con el nombre de Teoría de las Muestras.

Esta parte del Método Estadístico nos enseña a resolver, entre muchos otros importantes problemas, el que consiste, por

ejemplo, en determinar la significación de la diferencia de las medias de dos muestras aleatorias ; es decir, a determinar, en términos de probabilidad y de conformidad a determinadas hipótesis de trabajo, como la de la normalidad, si las muestras pueden o no provenir de un mismo colectivo. Otro de los problemas habituales que nos resuelve el muestreo es el de estimar los parámetros del colectivo, a partir de los estadísticos de las muestras, estimación que se hace en términos de probabilidad y con un predeterminado margen de error.

Fácil nos sería multiplicar casi hasta el infinito los problemas que nos resuelve y de los que se ocupa el Método Estadístico pero, hasta la publicación de la Memoria de Gini (Ref. 11) : *Il Concetto di Transvariazione e le sue Prime Applicazioni*, no encontramos resuelto por el Método Estadístico, entre otros análogos, el problema que se plantea de determinar el grado de significación del signo o de la intensidad de la diferencia que presentan dos observaciones tomadas al azar, una de cada colectivo, cuando se conoce la ley de distribución o el conjunto de valores de los elementos de cada colectivo, cuyas respectivas medias o medianas presentan una diferencia de signo contrario al de las dos observaciones dadas. Para la solución de este problema no es suficiente, la determinación de la diferencia de las medias, ni tampoco sus coeficientes de variabilidad, ni aún se resuelve recurriendo simultáneamente a ambos parámetros, como una cantidad de problemas en el muestreo lo permite. Es necesario agregar al conocimiento de la diferencia de las medias y de sus respectivos coeficientes de variabilidad, el conocimiento de las formas de sus correspondientes curvas de distribución.

Los parámetros de la transvariación sintetizan simultáneamente todas estas exigencias. Ellos son, entre otros, el campo de transvariación, la probabilidad de transvariación, la intensidad de transvariación, el área y el espacio de transvariación, etc., todos los cuales serán objeto de nuestro estudio.

Gini, con su memoria anteriormente mencionada, publicada en 1916, introduce por primera vez en la Metodología Estadística este importante y fecundo concepto, llenando en verdad, un vacío que más que nadie lo sentían los estudiosos de la Biometría y en particular de la Antropometría. Al respecto resulta ilustrativo mencionar a H. de Vries (Gini, Ref. 11) quien en su trabajo *Die*

Mutationstheorie, publicado en 1901, llama variabilidad transgresiva al hecho que cuando del estudio que se realiza de dos especies biológicas con respecto a un determinado atributo, nos encontramos con que el valor del atributo considerado, en dos ejemplares pertenecientes uno a cada especie, presentan entre si una diferencia de signo contrario al signo que presenta la diferencia de los respectivos valores típicos de ambas especies.

Evidentemente esta expresión de variabilidad transgresiva debida a de Vries tiene su importancia desde que significa un esfuerzo por llenar un vacío conceptual profundamente sentido por los biólogos ; pero de ella a la expresión analítica y cuantitativa que nos dé la medida de esta variabilidad, fluctuación transgresiva o transgresión, como también se acostumbra a llamarla, media una considerable distancia que es recorrida por la Memoria de Gini y luego, poco a poco, ampliada por el mismo autor y por su escuela, con fecundidad de resultados tanto en el campo teórico como en el de las aplicaciones.

Análoga situación se presentaba en sus orígenes, con el concepto de probabilidad. De la vaga, imprecisa, intuitiva y no mensurable expresión de es poco probable, es probable o es muy probable, de que se presente un acontecimiento previsto, a la expresión cuantitativa o sea a la medida de este concepto y a sus propiedades y a la estructuración de la Teoría Matemática de la Probabilidad, mediaba una distancia que fué llenada con el trabajo incansable de algunos siglos.

En 1949, con motivo de consideraciones análogas sobre la aparente dualidad de las direcciones cualitativa y cuantitativa en la investigación científica, en cada dominio del conocimiento, citábamos a Alexis Carrel ⁽¹⁾ cuando dice : « Hay una desigualdad extraña entre las ciencias de las materias inertes y las de los seres « vivos. La astronomía, la mecánica y la física tienen por base, « conceptos susceptibles de expresarse, de manera concisa y elegante, « en lenguaje matemático. Ellas han dado al universo las líneas « armoniosas de los monumentos de la Grecia antigua. Ellas lo

(1) CARREL, ALEXIS : *L'Homme cet Inconnu*, vertida al español por Ed. Joaquín Gil, Buenos Aires, bajo el título *La Incógnita del hombre*. La referencia corresponde al trabajo del A. : *La función de demanda : Su elasticidad - Flexibilidad del precio* - Op. cit. pág. 390.

« envuelven con la brillante red de sus cálculos y de sus hipótesis. « Persiguen la realidad más allá de las formas habituales del pensamiento hasta indecibles abstracciones hechas solamente de « ecuaciones y de símbolos. No es lo mismo para las ciencias biológicas. Los que estudian los fenómenos de la vida están como « perdidos en un matorral inextricable, en medio de una selva « mágica, cuyos árboles innumerables cambiarían sin cesar de « lugar y de forma. Se sienten agobiados, bajo un montón de hechos, que llegan a describir, pero que no son capaces de definir « con fórmulas algebraicas », para luego responder por nuestra cuenta, inspirándonos en un concepto de nuestro respetado jefe de la incipiente escuela argentina de Estadística, Carlos E. Dieulefait : « otro es el camino que han seguido las ciencias sociales, « pudiéndose apreciar sus progresos en la medida en que fueron « pasando de las nociones cualitativas a las nociones cuantitativas. « Más aún, se llegó a afirmar que la cantidad de matemática introducida en una disciplina, condiciona su progreso y su jerarquía científica »⁽²⁾. En esta dirección han orientado sus trabajos muchos matemáticos, filósofos, sociólogos, estadísticos y biólogos contemporáneos como Quetelet, Galton, Weldon, Karl Pearson, Gosset, R.A. Fisher, Gini, Boldrini y otros, sin olvidar al Rdo. Susmilch.

Lo anteriormente expresado nos permite suscribir la afirmación de Paul Painlevé, cuando dice en el prefacio de la traducción francesa de la obra de W. Stanley Jevons, « La Teoría de la Economía Política : « Es la marcha natural de las ciencias de evolucionar del estado cualitativo y descriptivo al estado cuantitativo y causal », con nuestras pequeñas reservas sobre la causalidad que sostiene, pues no admitimos en ningún dominio del conocimiento científico, la causalidad rigurosa o determinismo absoluto, al que respondería la definición de causalidad debida a Hume, cuando expresa⁽³⁾ : « si a un suceso A sigue constantemente, « indefectiblemente, un suceso B y B sólo se manifiesta después

(2) DAGUM, CAMILO : *La Función de Demanda : Su Elasticidad - Flexibilidad del Precio* - Op. Cit., pág. 391.

(3) JORDAN, P. : *La Mecánica Cuantista y los Problemas Fundamentales de la Biología y la Psicología* - Revista de Occidente - Madrid - Año X Nº CXIV - Diciembre 1932, pág. 225.

« de haber aparecido A-mientras que A puede aparecer sin pre-
cederle B-, entonces decimos que A es « causa » de B y B « efe-
cto » de A »; tampoco suscribimos el indeterminismo absoluto
o acausalidad, ni aún en el dominio de las ciencias sociales.

Para quien haya incursionado por la historia del Método Estadístico y en particular por los dominios actuales del mismo, con sus conclusiones en términos de probabilidad y a partir de determinadas hipótesis de trabajo, con el estado actual de desarrollo de los estudios de las variables aleatorias y fundamentalmente de las funciones aleatorias o procesos estocásticos, no tendrá reparo en suscribir la feliz expresión de Bell ⁽⁴⁾, cuando dice, « El moderno Método Estadístico es la matemática social por excelencia » « y cuando luego agrega, « Para comprender y analizar las reacciones de masas, sean estas de átomos o de seres humanos — de seres vivos diríamos más acertadamente —, se necesita dominar los modernos métodos estadísticos ».

Para un juicio sumamente gráfico sobre la ubicación de las ciencias físicas y las ciencias sociales entre los extremos de causalidad e indeterminismo, de donde respectivamente fueron sacadas por el esfuerzo creador de los hombres de ciencia de nuestro siglo, nos remitimos nuevamente a Bell. Este autor expresa ⁽⁵⁾: « Al irse haciendo la física moderna cada vez más ciertamente incierta, « las matemáticas de las probabilidades han aumentado su importancia científica. La mecánica newtoniana se aplica a una ciencia completamente determinada en que las ecuaciones diferenciales contienen la historia futura de un universo mecánicamente determinado. Para la interpretación científica de los experimentos de laboratorio, particularmente para los de física atómica, el método estrictamente mecánico de Newton, Lagrange, Laplace y sus sucesores, que se originó en la mecánica de Galileo y en la astronomía dinámica, ya no es adecuado y lo suplen cada vez más las matemáticas de la estadística y de la probabilidad ». Posteriormente agrega, « sin embargo, el número todavía gobierna el universo de las ciencias exactas, y

(4) BELL, ERIC TEMPLE: *Historia de las Matemáticas*. Edición traducida del inglés por el Fondo de Cultura Económica - México - 1949 - pág. 596.

(5) BELL, E.T. - op. cit., págs. 164 y 596.

« desde los comienzos del siglo XX el número ha empezado a penetrar en las ciencias menos exactas. Siempre que se aplica el método estadístico, la probabilidad matemática está implícita, y esta probabilidad se mide con números. Si la aplicación de la probabilidad matemática a las ciencias físicas las suavizó haciendo de ellas algo menos determinantemente rígido de lo que eran en el siglo XIX, al aplicarla análogamente a las ciencias sociales, las endureció con un toque de determinismo ».

La Teoría de la Transvariación, debida a Gini, tiene una breve pero fecunda historia. Por obra de la escuela italiana de estadística se encuentra en su actual grado de desarrollo, contribuyendo a la estructuración de la teoría y a la deducción de la varianza en el muestreo de la probabilidad e intensidad de transvariación, además de su creador, Castellano, Ottaviani, Livada, Sonnino, de Lucía, Viterbo, Benedetti y Herzel. En el campo de las aplicaciones, son de particular interés los trabajos de Boldrini, donde estudia los caracteres sexuales secundarios; otros autores como Giglio, Miani Calabrese y Filipponi, se han ocupado también de las aplicaciones de la transvariación en el dominio de las ciencias biológicas.

Con relación a los resultados obtenidos por Boldrini en el estudio de 50 series clasificadas conforme a sus atributos cuantitativos, para cada sexo, con el concurso metodológico de la transvariación, expresa Gini ⁽⁶⁾ « estas cifras han servido para aclarar múltiples cuestiones relativas a las diferencias secundarias entre el hombre y la mujer y han conducido a Boldrini a formular algunas leyes de notable interés alrededor de la sexualidad de los caracteres humanos ».

Contando con tan valiosas fuentes, desarrollamos a continuación, en forma sistemática, la Teoría de la Transvariación. En cuanto a las aplicaciones, que con tan fecundo resultado han sido llevadas al campo biológico, iniciamos aquí, entre otras, una serie de aplicaciones en el campo económico. Objeto posterior de ellas será un problema que desde hace algún tiempo es propósito del A. encarar su estudio cuantitativo. Se trata del estudio de las series de precios clasificadas en dos grandes grupos: precios agrí-

(6) GINI CORRADO, *Memorie di metodologia Statistica* - Vol. I - Ed. Giufre - Milano - 1939 - Pag. 490.

colas y precios industriales, en relación a su comportamiento a través de los procesos cíclicos del sistema económico. El argumento es de vital importancia para los países agrícolas, que coinciden plenamente con los económicamente subdesarrollados y para quienes no rigen las pretendidas ventajas de la división internacional del trabajo, cuya aplicación redundaría en beneficio de los países económicamente desarrollados, con la paradójica consecuencia, puesta de manifiesto por la teoría del desarrollo económico, que son los países subdesarrollados los que contribuyen a financiar el desarrollo de los países económicamente desarrollados.

Inmediatamente después de la primera guerra mundial, cobró singular importancia el problema de los precios y de los abastecimientos de las materias primas, problema que fué objeto de meditado estudio por Gini, por especial encargo de la entonces Sociedad de las Naciones. Este problema, desde entonces, no perdió actualidad. Acompaña a la gran crisis, dándole la contribución de su propio angustioso problema. En el orden internacional se suceden las conferencias entre los países interesados, entre las que destacamos, por su importancia, la que trató el problema del trigo, mientras en el orden interno de los países agrícolas se suceden las medidas, no siempre adecuadas, para resolver la situación. Vemos así intervenir decididamente al Estado con medidas tales como la adquisición de las cosechas a un precio dado, medida social con la que tiende a sostener los productores y con ellos el nivel de ocupación, frente a la brusca caída de los precios. Se crean una serie de juntas, encargadas de la regulación agrícola, con todo lo cual se inicia una etapa de decisiva intervención por parte del Estado, que caracteriza la vida económica contemporánea. El ideal de los clásicos y en particular de los fisiócratas, del estado gendarme, cuya misión la reducían a la función de policía, a la instrucción pública y a la construcción de obras públicas, pertenecen definitivamente al dominio de las especulaciones. La realidad de la vida económica contemporánea, su complejidad, la importancia del Estado en dicha vida, como elemento activo de la producción en el orden económico, de la distribución en el orden social, como consumidor y por sobre todo como poder político responsable de la realización de su programa de vida nacional, a través del equipo gobernante, exige su participación activa en el funcionamiento del sistema económico. La forma y

la intensidad de esta participación — decimos participación y no intervención y en ello va ya una toma de posiciones — es asunto que cae fuera de los límites de este estudio, pero diremos, sí, por considerar sumamente gráfica la figura, que concebimos al Estado casi como un virtuoso director de orquesta, austero, preciso y seguro, conduciendo armoniosamente el conjunto y con la participación responsable de todos sus integrantes. Hemos empleado deliberadamente el calificativo « casi » porque admitimos, dentro de ciertos límites condicionados por razones de orden político, económico y social, de la defensa nacional, la realización directa de un conjunto de actividades económicas.

Finalizada la segunda guerra mundial, vuelve nuevamente al primer plano de la consideración pública el problema de los precios de las materias primas con respecto a los de la industria. Particularmente en el seno de la O.E.A. (Organización de los Estados Americanos), se han desarrollado intensas conferencias de expertos económicos en busca de una solución que armonice con los intereses de todos.

Las aplicaciones en el campo económico que realizamos al final del presente tratamiento teórico, comprenden las importantes distribuciones de los índices de los obreros ocupados, horas-obrero trabajadas, poder adquisitivo del monto de los salarios pagados y volumen físico de la producción, correspondiente a 83 ramas de industria, abarcando las industrias extrativas, las manufactureras y las de electricidad y gas.

1.2 — NOTACIONES: Con Λ simbolizaremos en adelante un experimento aleatorio (abreviadamente e.a.), al cual asociamos una variable aleatoria (abreviadamente v.a.) que simbolizaremos con una letra griega minúscula (ξ, η, γ, \dots). Dicha v.a. puede asumir valores en el espacio R_k (k -dimensiones), siendo $k = 1, 2, \dots$. Con las últimas letras de nuestro abecedario (t, u, v, x, y) simbolizaremos una particular determinación en el muestreo, de la v.a. considerada en el e.a. dado. Con $f(x)$ simbolizaremos la densidad de probabilidad o función de densidad (abreviadamente f. de d.) y con $F(x)$ su correspondiente función de distribución o función de probabilidad acumulada (abreviadamente f. de p.a.), de la v.a. considerada, de la cual es x una particular determinación, es decir, su correspondiente variable estadística o variante.

En cada caso especificaremos el espacio en el cual estamos trabajando. Para el caso general en que la v.a. ξ asuma valores en el espacio R_k , corresponderá respectivamente, la siguiente notación para ξ , x , $f(x)$ y $F(x)$, considerando sus k dimensiones:

$$\xi = (\xi_1, \dots, \xi_k)$$

$$x = (x_1, \dots, x_k)$$

$$f(x) = f(x_1, \dots, x_k)$$

$$F(x) = P(\xi \leq x) = P(\xi_1 \leq x_1, \dots, \xi_k \leq x_k) = F(x_1, \dots, x_k)$$

En los capítulos que siguen, definiremos, para $k \geq 1$, los parámetros de la transvariación y deduciremos sus respectivas expresiones matemáticas. Para $k \geq 2$, partiremos de la hipótesis de v.a. dependientes, deduciendo como caso particular de ésta, los parámetros de la transvariación en la hipótesis de v.a. independientes. A su vez, como caso particular de cada una de estas hipótesis, trabajaremos con el supuesto de v.a. normalmente correlacionadas y v.a. independientes y normalmente distribuidas.

El estudio de la varianza y covarianza en el muestreo de los parámetros de la transvariación, será objeto de un capítulo especial.

2 - TRANSVARIACION ENTRE UNA DISTRIBUCION Y UNA CANTIDAD

2.1. INTRODUCCION

Sea Λ un e.a. al cual asociamos una v.a. ξ , que puede asumir valores en el espacio R_1 , cuya f. de d. es $f(x)$ y cuya f. de p.a. es:

$$P(\xi \leq x) = F(x) = \int_{-\infty}^x dF(t) = \begin{cases} \int_{-\infty}^x f(t) dt & \text{en el campo continuo} \\ \sum_{x_i \leq x} p_i & \text{en el campo discreto} \end{cases}$$

Introduciendo los parámetros λ y A , por ahora arbitrarios, nos definirá una transvariación entre la v.a. ξ y el parámetro A ,

con respecto al parámetro λ , en una realización determinada del e.a. Λ , cuando el signo de la diferencia entre el valor x de la v.a. ξ , en la realización del e.a. Λ , y el parámetro A , sea opuesto al signo de la diferencia entre los parámetros λ y A . O sea que, cada observación de la v.a. ξ , para la cual se verifique una diferencia $x - A$ de signo opuesto a la diferencia $\lambda - A$, nos dará, por definición, una transvariación. El valor absoluto de la potencia $(x - A)^r$ nos definirá su respectiva intensidad de orden r , de la cual obtenemos, para la intensidad de primer orden, el valor $|x - A|$, que nos define, al mismo tiempo, la medida de dicha transvariación.

Para todos los valores posibles de la v.a. ξ , en el espacio R_1 , para los cuales exista transvariación, la mayor diferencia $x - A$, en valor absoluto, nos definirá el campo de transvariación entre la v.a. ξ y el parámetro A , con respecto al parámetro λ .

Si nos encontramos con s casos para los cuales sea $x - a = 0$, siendo $\lambda \neq A$, convendremos en considerar la mitad de ellos como teniendo signo contrario al signo de la diferencia $\lambda - A$. Evidentemente esta convención es de aplicación cuando trabajamos con valores reales, como se da en el muestreo, pues estos constituyen una masa discreta de observaciones. Así también, es de aplicación cuando trabajamos con los valores teóricos, si su v.a. sigue una ley discreta de distribución.

Si en cambio es $\lambda = A$, sea que trabajemos con valores reales o teóricos y en este último supuesto, sea que su correspondiente v.a. obedezca a una ley de distribución discreta o continua, convendremos en considerar el número total de transvariaciones posibles como igual a la mitad del conjunto de observaciones y la intensidad de transvariación de orden r , para cada observación, convendremos en hacerla igual a

$$\frac{1}{2} |x - A|^r$$

2.2. TRANSVARIABILIDAD

Por definición, llamamos transvariabilidad entre la v.a. ξ y el parámetro A , con respecto al parámetro λ , a la probabilidad de que en una realización cualquiera del experimento aleatorio Λ ,

el valor observado de ξ asuma un valor x tal que la diferencia $x - A$ sea de signo opuesto al signo de la diferencia $\lambda - A$.

Lo anteriormente expresado equivale a decir que la transvariabilidad es igual a la probabilidad de que ξ asuma un valor menor de A si es $\lambda > A$ y a la inversa, si es $\lambda < A$.

Simbolizando con p la transvariabilidad, tenemos

$$p = \begin{cases} P(\xi \leq A) = \int_{-\infty}^A dF(x) = F(A) & \text{para } \lambda > A \\ P(\xi \geq A) = \int_A^{\infty} dF(x) = 1 - F(A) & \text{para } \lambda < A \end{cases} \quad (2.1.)$$

Evidentemente la transvariabilidad es igual al área comprendida entre la curva de probabilidad y la abscisa, en el intervalo $(-\infty, A)$, para el primer caso, y en el intervalo $(A, +\infty)$ para el segundo, en el supuesto de variable continua. Si trabajáramos con variables discretas, en lugar de área hablaríamos de frecuencia de transvariación y tal frecuencia nos definiría la transvariabilidad.

El valor de p no depende de los valores de λ y A independientemente considerados, sino de su ubicación relativa en el espacio R_1 , de la v.a. ξ , cuya f.d. es $f(x)$.

2.3. MAXIMO DE LA TRANSVARIABILIDAD

Teniendo presente la definición de transvariabilidad que nos conduce a la (2.1), definimos a continuación la transvariabilidad como función de un parámetro real t , tal que:

$$p(t) = \begin{cases} P(\xi \leq A + t) = F(A + t) & \text{para } \lambda > A + t \\ P(\xi \geq A + t) = 1 - F(A + t) & \text{para } \lambda < A + t \end{cases} \quad (2.2.)$$

de donde, recordando las propiedades de toda f . de $p.a.$, deducimos inmediatamente que $p(t)$ asume un valor máximo para $t = \lambda - A$, siempre que λ sea un punto de continuidad de $F(x)$; de no serlo, la condición de máximo para $p(t)$ viene satisfecha para cualquier valor de t , en el intervalo semiabierto por la derecha, en el primer caso y semiabierto por la izquierda, en el segundo,

dado por los dos puntos sucesivos de continuidad de ξ que contengan el punto de abscisa λ , en virtud del comportamiento de toda distribución discreta de probabilidad.

En consecuencia, el máximo de $\phi(t)$, que simbolizamos por ϕ_M es:

$$\phi_M = \lim_{t \rightarrow \lambda - A} \phi(t) = \begin{cases} \int_{-\infty}^{\lambda} dF(x) = F(\lambda) & \text{para } A < \lambda \\ 1 - \int_{-\infty}^{\lambda} dF(x) = 1 - F(\lambda) & \text{para } A > \lambda \end{cases} \quad (2.3.)$$

siendo $\phi_M \neq \frac{1}{2}$, siempre que λ no sea el valor mediano de la función de distribución.

En el campo de las aplicaciones, con vistas ya al resultado útil de la teoría, dado por la última etapa del método, que llamamos Inferencia Estadística, trabajaremos con λ igual al valor mediano, por verificarse, precisamente en este caso que

$$\phi_M = F(\lambda) = 1 - F(\lambda) = \frac{1}{2} \quad (2.4.)$$

El máximo de la transvariabilidad depende en consecuencia, exclusivamente, de λ , o en otras palabras, es función de λ , tendiendo a ser iguales ambas componentes de la fórmula (2.3), a medida que λ tiende al valor mediano de la distribución. En ningún caso su valor podrá ser superior a la unidad, lo que se desprende del hecho que la medida de la transvariabilidad se expresa en términos de probabilidad.

El mínimo de la transvariabilidad es función, exclusivamente, del parámetro A , teniendo por límite cero, cuando A cae fuera del espacio R_1 , en que viene definida ξ .

En efecto:

$$\lim_{A \rightarrow \infty} \phi = \begin{cases} \lim_{A \rightarrow -\infty} F(A) = 0 & \text{para cuando } \lambda > A \\ 1 - \lim_{A \rightarrow +\infty} F(A) = 0 & \text{para cuando } \lambda < A \end{cases} \quad (2.5.)$$

En la hipótesis que λ sea igual al valor mediano, deducimos, a partir de (2.4.) y (2.5.):

$$0 \leq p \leq \frac{1}{2} \quad (2.6.)$$

La transvariabilidad es una medida de la transvariación, pero, las consideraciones precedentemente realizadas, ponen de manifiesto algunas deficiencias que la hacen inadecuada para cumplir tal objetivo. La necesidad de dar una medida de la transvariación que pueda asumir cualquier valor dentro de un intervalo invariable, cualquiera sea el valor de λ , unida a la conveniencia que dicho intervalo sea el intervalo $[0,1]$, condujo a Gini (Ref. 11) a introducir el concepto de probabilidad de transvariación.

2.4. PROBABILIDAD DE TRANSVARIACION

El cociente entre la transvariabilidad y el máximo que ella puede asumir, es, por definición, la probabilidad de transvariación.

Simbolizándolo con P , tenemos:

$$P = \frac{p}{p_M} = \begin{cases} \frac{P(\xi \leq A)}{P(\xi \leq \lambda)} = \frac{F(A)}{F(\lambda)} & \text{para } \lambda \geq A \\ \frac{P(\xi \geq A)}{P(\xi \geq \lambda)} = \frac{1 - F(A)}{1 - F(\lambda)} & \text{para } \lambda \leq A \end{cases} \quad (2.7.)$$

Entre las características que se desprenden de la probabilidad de transvariación, en relación con la primera de (2.7.), señalamos:

1) es igual a cero para $A \rightarrow -\infty$ y, en general, cuando es $P(\xi \leq A) = 0$, siempre que sea $F(\lambda) \neq 0$;

2) si se verifica que también es $F(\lambda) = 0$, convenimos en definir igualmente como nula la probabilidad de transvariación;

3) es igual a la unidad cuando es $A = \lambda$ y $F(\lambda) \neq 0$;

4) es también igual a la unidad, cuando es $P(\xi \leq A) = 1$, siendo A , en tal caso, mayor que cualquier valor que pueda asumir ξ , en la realización del experimento aleatorio Λ ;

5) dado λ , es función no decreciente de A y dado A , es función no creciente de λ ;

6) es igual al doble de la transvariabilidad, cuando λ es igual al valor mediano;

7) su campo de variabilidad está dado siempre por el intervalo $[0,1]$.

Con las necesarias modificaciones del caso, impuestas por la condición $\lambda \leq A$, deducimos las características análogas para la segunda fórmula de (2.7).

2.5. MOMENTO DE TRANSVARIACION

La suma de los valores absolutos de las intensidades de orden r de todas las transvariaciones posibles de la v.a. ξ y el parámetro A , con respecto al parámetro λ , nos define el momento de transvariación de orden r .

Simbolizando por m_r dicho momento, tenemos:

$$m_r = \begin{cases} \int_{-\infty}^A (A-x)^r dF(x) & \text{para } \lambda > A \\ \int_A^{\infty} (x-A)^r dF(x) & \text{para } \lambda < A \end{cases} \quad (2.8.)$$

siendo para $r = 0$, $m_0 = p =$ transvariabilidad.

A los efectos de analizar el comportamiento de m_r como función de r , y teniendo presente una de las propiedades de las integrales definidas, tenemos, respectivamente, para $\lambda > A$ y $\lambda < A$

$$\frac{d m_r}{d r} = \begin{cases} \int_{-\infty}^A (A-x)^r \log(A-x) dF(x) = \int_{-\infty}^{A-1} (A-x)^r \log(A-x) dF(x) + \\ \quad + \int_{A-1}^A (A-x)^r \log(A-x) dF(x) \\ \int_A^{\infty} (x-A)^r \log(x-A) dF(x) = \int_{A+1}^{\infty} (x-A)^r \log(x-A) dF(x) + \\ \quad + \int_A^{A+1} (x-A)^r \log(x-A) dF(x) \end{cases} \quad (2.9)$$

Observando los dos sumandos de los últimos miembros de las dos igualdades precedentes, vemos que nada podemos afirmar, en forma general, sobre el signo de la derivada del momento de transvariación. En efecto, al ser los primeros sumandos de los dos últimos miembros que estamos considerando, permanentemente no negativos, y los dos últimos sumandos permanentemente no positivos, el signo de $\frac{d m_r}{d r}$ depende del sumando de mayor valor absoluto, el que varía no sólo para cada distribución en particular, sino que, más aún, varía dentro de una misma distribución, con el variar de r y también para distintos valores de A . En consecuencia, del análisis de la derivada del momento de transvariación, nos encontramos con que m_r puede ser: una función constante de r , creciente, decreciente o presentar algún mínimo, por ser.

$$\frac{d^2 m_r}{d r^2} \geq 0$$

Se puede, sin inconvenientes, elaborar ejemplos ilustrativos de cada una de estas situaciones. Veamos uno.

Tengamos la siguiente distribución de frecuencias:

x	f_x	
1	2	con media aritmética: $\bar{x} = 3,50$
2	10	
5	10	con valor mediano M que lo hacemos igual a 3,50, media
6	2	
	$\frac{2}{24}$	aritmética del intervalo (2,5).

Hacemos $\lambda = 3,50$ y $A = 2,10$, con lo que tenemos, para el momento de transvariación:

$$m_r = \frac{\sum_{x=1}^2 (2,10 - x)^r f_x}{24} = \left\{ \begin{array}{l} \text{para } r = 0: m_0 = 0,50 \\ \text{» } r = 1: m_1 = 0,1333 \\ \text{» } r = 2: m_2 = 0,1050 \\ \text{» } r = 3: m_3 = 0,1113 \\ \text{» } r = 4: m_4 = 0,12205 \end{array} \right.$$

$$= \frac{2(1,10)^r + 10(0,10)^r}{24}$$

de donde, a partir de

$$m_r = \frac{2 (1,10)^r + 10 (0,10)^r}{24}$$

deducimos las siguientes conclusiones, para r no negativo :

- 1) m_r es una función continua de r ;
- 2) m_r presenta un mínimo absoluto, en el campo real positivo, para $r = 2$; ($r = 1,999.284$) ;
- 3) para $r = 0$ es $m_0 = 0,50$;
- 4) para $r \rightarrow \infty$ es $\lim m_r = +\infty$

Si hacemos ahora $A = 3$, tenemos :

$$m_r = \frac{\sum_{x=1}^2 (3-x)^r f_x}{24} = \frac{2^{r+1} + 10}{24} \left\{ \begin{array}{l} \text{Para } r = 0 : m_0 = 0,50 \\ \text{» } = 1 : m_1 = 0,5833 \\ \text{» } = 2 : m_2 = 0,7500 \\ \text{» } = 3 : m_3 = 1,0833 \\ \text{» } = 4 : m_4 = 1,2500 \end{array} \right.$$

de donde deducimos, a partir de

$$m_r = \frac{10 + 2^{r+1}}{24}$$

- 1) m_r es una función monótonamente creciente de r ;
- 2) $m_r = 0,50$ para $r = 0$;
- 3) $\lim_{r \rightarrow +\infty} m_r = +\infty$

Para $A = 2$, tenemos para m_r , en virtud de lo convenido en (2.1) :

- 1) m_r es una función constante e igual a $\frac{1}{12}$, para $r > 0$;
- 2) $m_0 = 0,29166$, presentando m_r , para $r = 0$, un punto de discontinuidad.

Llamando $\beta_{r,A}$ al momento absoluto de orden r de la v.a. ξ con respecto al origen arbitrario A , tenemos :

$$\beta_{r,A} = E (|\xi - A|^r) \geq m_r \quad (2.10)$$

En efecto :

$$\beta_{r,A} = E(|\xi - A|)^r = \left\{ \begin{array}{l} \int_{-\infty}^A (A-x)^r dF(x) + \int_A^{\infty} (x-A)^r dF(x) = m_r + \\ \quad + \int_A^{\infty} (x-A)^r dF(x) \text{ para } \lambda > A \\ \int_A^{\infty} (x-A)^r dF(x) + \int_{-\infty}^A (A-x)^r dF(x) = m_r + \\ \quad + \int_{-\infty}^A (A-x)^r dF(x) \text{ para } \lambda < A \end{array} \right. \quad (2.11)$$

Luego, el momento de transvariación es una parte del momento absoluto con respecto al origen arbitrario A , llegando, en el límite, a ser iguales, cuando sea nula la integral del último miembro de cada una de las igualdades precedentes.

2.6. MAXIMO DEL MOMENTO DE TRANSVARIACION

Definimos a continuación el momento de transvariación como función del parámetro real t . A partir de (2.8), tenemos,

$$m_r(t) = \left\{ \begin{array}{l} \int_{-\infty}^{A+t} (A+t-x)^r dF(x) \text{ para } \lambda > A+t \\ \int_{A+t}^{\infty} (x-A-t)^r dF(x) \text{ para } \lambda < A+t \end{array} \right. \quad (2.12)$$

La expresión subintegral de la primera fórmula precedente es una función positiva y monótonamente creciente con respecto al parámetro real t . En consecuencia, su integral, observa análogo comportamiento, alcanzando su máximo con el máximo valor que puede asumir t , que en nuestro caso es $t = \lambda - A$.

La segunda expresión subintegral de (2.12), es una función positiva y monótonamente decreciente con respecto al parámetro real t en el intervalo $(\lambda - A, \infty)$.

En consecuencia, su integral alcanza su máximo para $t = \lambda - A$, que es el límite inferior del intervalo de variabilidad de t , que resulta de la definición de transvariación y de los parámetros de la transvariación, de la cual partimos.

Simbolizando por $m_{r,M}$ el máximo del momento de transvariación $m_r(t)$ tenemos:

$$m_{r,M} = \begin{cases} \int_{-\infty}^{\lambda} (\lambda - x)^r dF(x) & \text{para } \lambda > A \\ \int_A^{\infty} (x - \lambda)^r dF(x) & \text{para } \lambda < A \end{cases} \quad (2.13)$$

Luego, el máximo del momento de transvariación de la *v.a.* ξ y el parámetro A , con respecto al parámetro λ , se obtiene sustituyendo en (2.8) A por λ , siendo la primera de (2.13) el máximo de m_r , si es $\lambda > A$ y la segunda, si es $\lambda < A$.

En el campo de las aplicaciones, trabajaremos con λ igual a la media aritmética, que es el valor más apropiado para la extracción de conclusiones en cada caso concreto, por las siguientes razones:

1) en general, es el valor más representativo de la tendencia central de una distribución, salvo que la distribución considerada siga una ley, por ejemplo, definidamente geométrica o armónica, caso en que ella estaría mejor representada, respectivamente, por la media geométrica o la media armónica;

2) Teniendo presente el objetivo que toda teoría debe lograr para alcanzar la plenitud de su vigencia, cual es la posibilidad de su aplicación a hechos concretos de la realidad a través de su proceso de explicación, deducción de leyes, interpretación y previsión para el futuro, situación ideal en que una teoría conquista la « maternidad », nosotros trabajaremos con λ igual a la media

aritmética, la cual simbolizamos con m . Para $\lambda = m$ y $r = 1$, se verifica :

$$m_{1;M} = \int_{-\infty}^m (m - x) dF(x) = \int_m^{\infty} (x - m) dF(x) = \frac{1}{2} \beta_1 \quad (2.14)$$

cualquiera sea la ley de distribución de la v.a. ξ , en virtud de una de las propiedades de la media aritmética, siendo $\beta_1 = E(|\xi - m|)$, que nos define el desvío medio de la v.a. ξ ;

3) si la v.a. ξ sigue una ley de distribución simétrica, tendremos, para cualquier valor de r :

$$\begin{aligned} m_{r;M} &= \int_{-\infty}^m (m - x)^r dF(x) = \\ &= \int_m^{\infty} (x - m)^r dF(x) = \frac{1}{2} E(|\xi - m|^r) \end{aligned} \quad (2.15)$$

siendo también, m , por hipótesis de distribución simétrica, igual al valor mediano. $E(|\xi - m|^r)$ nos define el momento absoluto de orden r de la variable desvío. Simbolizándolo por β_r , tenemos :

$$\beta_r = E(|\xi - m|^r) \quad (2.16)$$

4) por tener la media aritmética una expresión analítica, lo que permite someterla a las operaciones del cálculo diferencial e integral.

El máximo del momento de transvariación es función exclusiva del parámetro λ . Para $r > 0$ se verifica, cuando λ asume valores en el intervalo :

$$-\infty < \lambda < +\infty$$

que el máximo del momento de transvariación varía en el intervalo $0 \leq m_{r;M} < \infty$ si es $\lambda > A$, y en el intervalo

$$-\infty < m_{r;M} \leq 0$$

si es $\lambda < A$.

2.7. INTENSIDAD DE TRANSVARIACION

El momento de transvariación, puede ser también considerado como una medida de la transvariación, pero, aisladamente, es de escasa utilidad para cumplir con este objetivo, como se desprende del análisis hecho en (2.5). La necesidad de dar una medida que asuma valores en el intervalo $[0,1]$, condujo a Gini (Ref. 11) como ya lo había hecho con la probabilidad de transvariación, a introducir el concepto de intensidad de transvariación.

Definimos intensidad de transvariación de orden r , a la raíz r -ésima del cociente entre el momento de transvariación de orden r y su valor máximo.

La intensidad de transvariación de orden r , para $r > 1$, ha sido introducida por L. de Lucía (Ref. 9).

Simbolizando con I , a la intensidad de transvariación, y haciendo $m_{(r)} = (m_r)^{\frac{1}{r}}$; $m_{(r,M)} = (m_{r,M})^{\frac{1}{r}}$, tenemos:

$$= \frac{m_{(r)}}{m_{(r,M)}} = \begin{cases} \frac{\left(\int_{-\infty}^A (A-x)^r dF(x) \right)^{\frac{1}{r}}}{\left(\int_{-\infty}^{\lambda} (\lambda-x)^r dF(x) \right)^{\frac{1}{r}}} & \text{para } \lambda \geq A \\ \frac{\left(\int_A^{\infty} (x-A)^r dF(x) \right)^{\frac{1}{r}}}{\left(\int_{\lambda}^{\infty} (x-\lambda)^r dF(x) \right)^{\frac{1}{r}}} & \text{para } \lambda \leq A \end{cases} \quad (2.17)$$

Señalamos algunas de las características de (2.17), en la hipótesis $\lambda \geq A$. Para la hipótesis $\lambda < A$ se deducen, con las necesarias modificaciones del caso, impuesto por los distintos límites de la integral, análogas conclusiones.

Dichas características son:

1) I_r es igual a cero cuando $A \rightarrow -\infty$ y, en general, es nula cuando $P(\xi \leq A) = 0$, siempre que $F(\lambda) \neq 0$;

2) Convenimos en definir como nula la intensidad de trans-
variación cuando es, simultáneamente, $P(\xi \leq A) = 0$ y $F(\lambda) = 0$;

3) es igual a uno cuando es $A = \lambda$ y $F(\lambda) \neq 0$;

4) dado λ , I_r es función monótonamente creciente de A ,
en el intervalo semiabierto $(-\infty, \lambda]$ y dado A , es función
monótonamente decreciente de λ , en el intervalo semiabierto
 $[A, +\infty)$;

5) es :

$$I_r = \frac{[E(|\xi - A|)^r]^{\frac{1}{r}}}{[E(|\xi - \lambda|)^r]^{\frac{1}{r}}} = \frac{\beta_{r,A}^{\frac{1}{r}}}{\beta_{r,\lambda}^{\frac{1}{r}}} \quad \text{en virtud de} \quad (2.10)$$

cuando A es mayor que el mayor valor que puede asumir la v.a. ξ
en la realización del e.a. Λ ;

6) es :

$$I_r = \frac{\frac{1}{m_r^{\frac{1}{r}}}}{\beta_{r,\lambda}^{\frac{1}{r}}}$$

cuando solamente es λ mayor que el mayor valor que puede
asumir la v.a. ξ ;

7) para $\lambda = m$, es :

$$I_r = \frac{(2 m_r)^{\frac{1}{r}}}{\beta_r^{\frac{1}{r}}}$$

cuando la v.a. ξ se distribuye en forma simétrica. β_r está dado
por (2.16). Para $r = 2$ es.

$$I_2 = \frac{(2 m_2)^{\frac{1}{2}}}{\beta_2^{\frac{1}{2}}} = \frac{2 m_2^{\frac{1}{2}}}{\sigma_\xi}$$

8) cualquiera sea la ley de distribución de la v.a. ξ , es, para $r=1$ y $\lambda=m$, en virtud de (2.14):

$$I_1 = \frac{2 m_1}{\beta_1}$$

9) I_r asume valores en el intervalo $[0,1]$.

2.8. TRANSVARIACION EN LA HIPOTESIS DE DISTRIBUCION UNIFORME

Sea el e.a. Λ , al cual asociamos la v.a. ξ , que puede asumir valores en el intervalo $[0, 2b]$, cuyo f. de d. es

$$f(x) = \begin{cases} \frac{1}{2b} & \text{en } [0, 2b] \\ 0 & \text{fuera de este intervalo} \end{cases} \quad (2.18)$$

y cuya f. de p.a. es

$$F(x) = \begin{cases} 0 & \text{para } x \leq 0 \\ \frac{x}{2b} & \text{para } 0 < x < 2b \\ 1 & \text{para } x \geq 2b \end{cases} \quad (2.19)$$

Toda distribución uniforme de probabilidad, definida en cualquier intervalo, puede ser expresada, mediante una adecuada sustitución de variable, con una distribución uniforme definida en el intervalo $[0,1]$ o en el dado precedentemente, con el cual trabajamos ahora.

Para la determinación de la probabilidad de transvariación, trabajamos con el valor mediano que simbolizamos con M_a y para la intensidad de transvariación con la media aritmética m , siendo, en este caso particular, en virtud de la simetría de la distribución con respecto a m :

$$M_a = m = b$$

Para $0 < A < b$, tenemos respectivamente, en virtud de (2.8), (2.13), (2.16) y (2.19):

$$m_r = \frac{1}{2b} \int_0^A (A-x)^r dx = \frac{A^{r+1}}{2b(r+1)} \quad (2.20)$$

$$m_{r,M} = \frac{1}{2b} \int_0^b (b-x)^r dx = \frac{b^r}{2(r+1)} \quad (2.21)$$

$$\beta_r = E(|\xi - b|^r) = \frac{b^r}{r+1} = 2 m_{r,M} \quad (2.22)$$

siendo

$$m_{r,M} = \frac{1}{2} \beta_r$$

por tratarse de una distribución simétrica, caso estudiado en (2.15).

Para $r_1 = 1$, la (2.22) nos define el desvío medio y para $r = 2$ la varianza de la distribución, siendo, respectivamente:

$$\beta_1 = E(|\xi - b|) = \frac{b}{2}$$

$$\beta_2 = E(\xi - b)^2 = \frac{b^2}{3}$$

Para $r = 0$, la (2.20) y la (2.21) nos definen, respectivamente, la transvariabilidad y el máximo de la transvariabilidad de esta distribución. Por lo tanto

$$\phi = m_0 = \frac{A}{2b}$$

$$\phi_M = m_{0,M} = \frac{1}{2}$$

y, por definición de probabilidad de transvariación;

$$P = \frac{\phi}{\phi_M} = \frac{A}{b} \quad (2.23)$$

Por definición de intensidad de transvariación de orden r , tenemos, a partir de (2.17), (2.20), (2.21) y (2.22):

$$I_r = \frac{(2 m_r)^{\frac{1}{r}}}{\beta_r^{\frac{1}{r}}} = \frac{A}{b} \sqrt[r]{\frac{A}{b}} = P \sqrt[r]{P} \quad (2.24)$$

siendo, para $r = 1$

$$I_1 = \frac{A^2}{b^2}$$

Verificándose la desigualdad

$$P > I_r$$

para todo r real y positivo, siendo asimismo I_r una función continua, monótonamente creciente, en dicho campo de variabilidad. En todo el campo real, I_r presenta un punto de discontinuidad para $r = 0$.

Dado el parámetro arbitrario λ , siendo $0 < A < \lambda$, se deduce inmediatamente que

$$P = \frac{A}{\lambda} \quad (2.25)$$

siendo asimismo

$$I_r = \frac{A}{\lambda} \sqrt[r]{\frac{A}{\lambda}} = P \sqrt[r]{P} \quad (2.26)$$

Para $\lambda < A$, se deducen en forma inmediata los correspondientes parámetros de la transvariación.

2.9. TRANSVARIACION EN LA HIPOTESIS DE DISTRIBUCION DE POISSON.

Dado el e.a. Λ , al cual asociamos la v.a. ξ , definida en el espacio R_1 , dado por cero y el conjunto de los números naturales, su f. de d. es

$$P_x = \begin{cases} \frac{m^x}{x!} e^{-m} & \text{para todo } x \in R_1 \\ 0 & \text{cuando } x \notin R_1 \end{cases} \quad (2.27)$$

y su f. de p.a. es

$$F(x) = e^{-m} \sum_{t=0}^{[x]} \frac{m^t}{t!} \quad (2.28)$$

siendo $[x]$ el mayor entero contenido en x .

Para $A < \lambda$, siendo $A > 0$, tenemos

$$m_r = \int_0^A (A-x)^r dF(x) = e^{-m} \sum_{x=0}^{[A]} (A-x)^r \frac{m^x}{x!} \quad (2.29)$$

$$m_{r;M} = \int_0^\lambda (\lambda-x)^r dF(x) = e^{-m} \sum_{x=0}^{[\lambda]} (\lambda-x)^r \frac{m^x}{x!} \quad (2.30)$$

Por definición de intensidad de transvariación de orden r , es:

$$I_r = \frac{m_{(r)}}{m_{(r;M)}} = \frac{\left(\sum_{x=0}^{[A]} (A-x)^r \frac{m^x}{x!} \right)^{\frac{1}{r}}}{\left(\sum_{x=0}^{[\lambda]} (\lambda-x)^r \frac{m^x}{x!} \right)^{\frac{1}{r}}} \quad (2.31)$$

para $r = 1, 2, 3, \dots$

Para $r = 0$, la (2.29) y la (2.30), nos definen, respectivamente, la transvariabilidad y el máximo de la transvariabilidad entre la v.a. ξ , cuya f. de d. está dada por (2.27), y el parámetro A , con respecto al parámetro λ . Luego, por (2.7), la probabilidad de transvariación es

$$P = \frac{p}{p_M} = \frac{m_0}{m_{0;M}} = \frac{\sum_{x=0}^{[A]} \frac{m^x}{x!}}{\sum_{x=0}^{[\lambda]} \frac{m^x}{x!}} \quad (2.32)$$

Si A y λ son números pares, se demuestra ⁽⁷⁾ que

$$e^{-m} \sum_{x=0}^A \frac{m^x}{x!} = \frac{1}{\Gamma(A+1)} \quad (2.33)$$

$$\int_m^\infty x^A e^{-x} dx = \frac{1}{2^{A+1} \Gamma(A+1)} \int_{2m}^\infty x^A e^{-\frac{x}{2}} dx = 1 - K_{2A+2}(2m)$$

$$e^{-m} \sum_{x=0}^\lambda \frac{m^x}{x!} = \frac{1}{\Gamma(\lambda+1)} \int_m^\infty x^\lambda e^{-x} dx = 1 - K_{2\lambda+2}(2m) \quad (2.34)$$

donde

$$K_n(t) = P(\chi^2 \leq t) = \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \int_0^t x^{\frac{n}{2}-1} e^{-\frac{x}{2}} dx$$

nos define la f. de p.a. de la v.a. χ^2 de Pearson, con n grados de libertad.

Para A y λ pares, se obtiene para la probabilidad de transvariación, a partir de (2.33) y (2.34):

$$P = \frac{p}{p_M} = \frac{1 - K_{2A+2}(2m)}{1 - K_{2\lambda+2}(2m)} \quad (2.35)$$

cuyo valor calculamos inmediatamente por medio de la tabla de la distribución χ^2 .

Para $\lambda < A$, se deducen en forma inmediata los correspondientes parámetros de la transvariación.

Igualmente se pueden obtener dichos parámetros para cuando sea λ igual al valor mediano o a la media aritmética.

⁽⁷⁾ Para la demostración correspondiente, nos remitimos a: FERNANDEZ BAÑOS, OLEGARIO en *Tratado de Estadística*, Ed. Dossat, Madrid, 1945, Pág. 284 y DAGUM, CAMILO en *Distribución de Poisson*, en «Revista de Economía», del Banco de la Provincia de Córdoba, (Argentina), Año IV, N° 8 julio-diciembre de 1952, pág. 211.

2.10. TRANSVARIACION EN LA HIPOTESIS DE DISTRIBUCION NORMAL

Dado el e.a. Λ , al cual asociamos la v.a. ξ normal (m, σ) cuya f. de d. es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad (2.36)$$

definida en todo el campo real y simétrica con respecto al valor m de x , que es igual a la media aritmética y al valor mediano de la distribución.

Su f. de p.a. es

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-m)^2}{2\sigma^2}} dt \quad (2.37)$$

Para la v.a. reducida $\eta = \frac{\xi - m}{\sigma}$ es decir para η normal $(0,1)$, su f. de p.a. que simbolizamos $\Phi(x)$ es:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (2.38)$$

Para $\lambda > A$, es

$$m_r = \int_{-\infty}^A (A - x)^r dF(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^A (A - x)^r e^{-\frac{(x-m)^2}{2\sigma^2}} dx \quad (2.39)$$

$$m_{r,M} = \int_{-\infty}^{\lambda} (\lambda - x)^r dF(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\lambda} (\lambda - x)^r e^{-\frac{(x-m)^2}{2\sigma^2}} dx \quad (2.40)$$

El momento absoluto de orden r con respecto a la media m es

$$\beta_r = E(|\xi - m|^r) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (|x - m|^r) e^{-\frac{(x-m)^2}{2\sigma^2}} dx \quad (2.41)$$

Para encontrar el valor de m_r en (2.39), realizamos la sustitución $t = \frac{x-m}{\sigma}$ y luego de hacer $b = \frac{A-m}{\sigma}$, tenemos:

$$m_r = \frac{\sigma^r}{\sqrt{2\pi}} \int_{-\infty}^b (b-t)^r e^{-\frac{t^2}{2}} dt = \frac{\sigma^r \cdot r!}{\sqrt{2\pi}} \sum_{s=0}^r \frac{b^{r-s}}{s!(r-s)!} \int_{-b}^{\infty} t^s e^{-\frac{t^2}{2}} dt$$

Separando en dos grupos los sumandos del último miembro, atendiendo al carácter par o impar de s , tenemos:

$$\begin{aligned} m_r = b^r \sigma^r \Phi(b) + \frac{\sigma^r}{\sqrt{2\pi}} \sum_{k=0}^{\alpha} b^{r-2k-1} \binom{r}{2k+1} \int_{-b}^{\infty} t^{2k+1} e^{-\frac{t^2}{2}} dt + \\ + \frac{\sigma^r}{\sqrt{2\pi}} \sum_{k=1}^{\left[\frac{r}{2}\right]} b^{r-2k} \binom{r}{2k} \int_{-b}^{\infty} t^{2k} e^{-\frac{t^2}{2}} dt \end{aligned} \quad (2.42)$$

siendo

$$\alpha = \begin{cases} \left[\frac{r}{2}\right] & \text{para } r \text{ impar} \\ \frac{r}{2} - 1 & \text{para } r \text{ par} \end{cases}$$

Simbolizando, respectivamente, con $G(2k+1; -b)$ y $G(2k; -b)$ cada una de las dos integrales de (2.42) e integrándolas por partes, resulta:

$$G(2k+1; -b) = \int_{-b}^{\infty} t^{2k+1} e^{-\frac{t^2}{2}} dt = b^{2k} e^{-\frac{b^2}{2}} + 2k \int_{-b}^{\infty} t^{2k-1} e^{-\frac{t^2}{2}} dt$$

$$G(2k; -b) = \int_{-b}^{\infty} t^{2k} e^{-\frac{t^2}{2}} dt = -b^{2k-1} e^{-\frac{b^2}{2}} + (2k-1) \int_{-b}^{\infty} t^{2k-2} e^{-\frac{t^2}{2}} dt$$

a partir de las cuales obtenemos las siguientes fórmulas de recurrencia que nos conducirán a la solución de cada una de dichas integrales

$$G(2k+1; -b) = b^{2k} e^{-\frac{b^2}{2}} + 2k G(2k-1; -b) \quad (2.43)$$

$$G(2k; -b) = -b^{2k-1} e^{-\frac{b^2}{2}} + (2k-1) G(2k-2; -b) \quad (2.44)$$

Siendo, respectivamente, para $k=0$

$$G(1; -b) = e^{-\frac{b^2}{2}} \quad (2.45)$$

$$G(0; -b) = \sqrt{2\pi} \Phi(b) \quad (2.46)$$

Para resolver $G(2k+1; -b)$, partimos de (2.43), teniendo presente (2.45), dando valores a k desde uno hasta k . Luego para $k=1$:

$$G(3; -b) = b^2 e^{-\frac{b^2}{2}} + 2 G(1; -b) = e^{-\frac{b^2}{2}} (b^2 + 2)$$

para $k=2$:

$$G(5; -b) = b^4 e^{-\frac{b^2}{2}} + 4 G(3; -b) = e^{-\frac{b^2}{2}} (b^4 + 4b^2 + 4 \cdot 2)$$

para el valor k :

$$\begin{aligned} G(2k+1; -b) &= e^{-\frac{b^2}{2}} \left[b^{2k} + \sum_{i=1}^k 2k(2k-2) \dots (2k-2i+2) b^{2k-2i} \right] = \\ &= e^{-\frac{b^2}{2}} \sum_{i=0}^k 2^i i! \binom{k}{i} b^{2k-2i} \end{aligned} \quad (2.47)$$

Procediendo en forma análoga con $G(2k; -b)$ tenemos, a partir de (2.44) y (2.46):

$$G(2k; -b) = \frac{(2k)!}{2^k k!} \sqrt{2\pi} \Phi(b) - \quad (2.48)$$

$$- e^{-\frac{b^2}{2}} \sum_{i=0}^{k-1} \frac{(2i)! \binom{2k}{2i}}{2^i i! \binom{k}{i}} b^{2k-2i-1}$$

A partir de (2.42), (2.47) y (2.48), tenemos para m_r :

$$\begin{aligned} m_r &= b^r \sigma^r \Phi(b) + \frac{\sigma^r e^{-\frac{b^2}{2}}}{\sqrt{2\pi}} \cdot \sum_{k=0}^{\alpha} b^{r-2k-1} \binom{r}{2k+1} \sum_{i=0}^k 2^i i! \binom{k}{i} b^{2k-2i} + \\ &+ \sigma^r \Phi(b) \sum_{k=1}^{\left[\frac{r}{2}\right]} b^{r-2k} \binom{r}{2k} \frac{2k!}{2^k k!} - \\ &- \frac{\sigma^r e^{-\frac{b^2}{2}}}{\sqrt{2\pi}} \sum_{k=1}^{\left[\frac{r}{2}\right]} b^{r-2k} \binom{r}{2k} \sum_{i=0}^{k-1} \frac{(2i)! \binom{2k}{2i}}{2^i i! \binom{k}{i}} b^{2k-2i-1} \end{aligned} \quad (2.49)$$

Si en (2.49) hacemos $b = \frac{\lambda - m}{\sigma}$, tenemos resuelto (2.40), o sea el máximo del momento de transvariación de orden r .

Para obtener el valor de β_r en (2.41), realizamos la sustitución $t = \frac{x - m}{\sigma}$, resultando

$$\beta_r = \frac{2\sigma^r}{\sqrt{2\pi}} \int_0^\infty t^r e^{-\frac{t^2}{2}} dt \quad (2.50)$$

integral que reducimos a una función gamma de Euler, mediante la sustitución $v = \frac{t^2}{2}$. Luego

$$\beta_r = \frac{2^{\frac{r}{2}} \sigma^r}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \quad (2.51)$$

siendo, para :

$$r = 2k : \beta_{2k} = \mu_{2k} = \frac{(2k)!}{2^k k!} \sigma^{2k} \quad (2.51.a)$$

$$r = 2k+1 : \beta_{2k+1} = \frac{2^{\frac{2k+1}{2}} k!}{\sqrt{\pi}} \sigma^{2k+1} \quad (2.51.b)$$

Mientras que

$$\mu_{2k+1} = 0 \quad \text{para } k = 0, 1, 2, \dots$$

dada la simetría de la distribución de la *v.a.* normal. El símbolo μ_r nos define el momento centrado de orden r .

Por definición de probabilidad de transvariación, tenemos :

$$P = \frac{\dot{p}}{\dot{p}_M} = \frac{m_0}{m_{0;M}} = \frac{F(A)}{F(\lambda)} \quad (2.52)$$

donde los valores $F(A)$ y $F(\lambda)$ los obtenemos a partir de la tabla de la distribución normal.

Por definición de intensidad de transvariación de orden r , es

$$I_r = \frac{m_{(r)}}{m_{(r;M)}} = \frac{m_r^{\frac{1}{r}}}{m_{r;M}^{\frac{1}{r}}} \quad (2.53)$$

donde m_r y $m_{r;M}$ los obtenemos a partir de (2.49), haciendo $b = \frac{A-m}{\sigma}$ para el primero y $b = \frac{\lambda-m}{\sigma}$ para el segundo.

Para el caso particular $\lambda = m$, siendo $F(m) = \frac{1}{2}$ y $m_{r,M} = \frac{1}{2} \beta_r$, tenemos, respectivamente, para (2.52) y (2.53):

$$P = 2 F(A) \quad (2.54)$$

$$I_r = \frac{(2 m_r)^{\frac{1}{r}}}{\beta_r^{\frac{1}{r}}} \quad (2.55)$$

donde β_r lo obtenemos a partir de (2.51.a) o (2.51.b) según sea r par o impar.

En la hipótesis $\lambda = m$, obtenemos para la intensidad de transvariación de orden uno, a partir de (2.49), (2.51) y (2.55):

$$I_1 = \frac{2 m_1}{\beta_1} = b \Phi(b) \sqrt{2\pi} + e^{-\frac{b^2}{2}}$$

Para $\lambda < A$, se deducen en forma inmediata los correspondientes parámetros de la transvariación, como así también para el caso particular $\lambda = m$, o sea $m < A$.

Las características de la probabilidad de transvariación y de la intensidad de transvariación, respectivamente dadas en 2.4 y en 2.7, se deducen en forma inmediata para cada uno de los casos particulares aquí considerados, a saber: 1) de distribución uniforme, 2) de distribución de Poisson y 3) de distribución normal, con sólo atender a las modalidades propias de cada una de estas distribuciones.

3. — TRANSVARIACION EN R_2

3.1. INTODUCCION

Dado un experimento aleatorio Λ , al cual asociamos las v.a. bidimensional

$$\xi = (\xi_1, \xi_2)$$

que puede asumir valores en el espacio R_2 , que salvo limitación expresa, se entenderá todo el plano real. Su f. de d. es $f(x, y)$, siendo $z = (x, y)$ una particular determinación de $\xi = (\xi_1, \xi_2)$.

La probabilidad de la relación $(\xi_1, \xi_2) \subset S$ es

$$P(S) = P(\xi \subset S) = P(\xi_1 \subset S_1, \xi_2 \subset S_2); \quad S = S_1 S_2$$

Siendo su f. de p.a. $F(x, y)$:

$$F(x, y) = P(\xi_1 \leq x, \xi_2 \leq y) = \int_s dP$$

y

$$\int_{R_2} dP = 1$$

Simbolizando por $F_1(x)$ la f. de p.a. de la distribución marginal de ξ_1 , y por $F_2(y)$ la f. de p.a. de ξ_2 , tenemos

$$F_1(x) = P(\xi_1 \leq x) = F(x, \infty)$$

$$F_2(y) = P(\xi_2 \leq y) = F(\infty, y)$$

Si la v.a. bidimensional ξ tiene una distribución continua de probabilidad, es

$$f(x, y) = \frac{\delta^2 F}{\delta x \delta y}$$

$$P(S) = \int_s f(x, y) dx dy = \int_{s_1} \int_{s_2} f(x, y) dx dy$$

$$F(x, y) = P(\xi_1 \leq x, \xi_2 \leq y) = \int_{-\infty}^y \int_{-\infty}^x f(x, y) dx dy$$

$$P(R_2) = F(\infty, \infty) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

$$F_1(x) = P(\xi_1 \leq x) = \int_{-\infty}^x f_1(t) dt$$

$$F_2(y) = P(\xi_2 \leq y) = \int_{-\infty}^y f_2(t) dt$$

siendo

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

para el caso discreto, tenemos

$$P(\xi_1 = x_i, \xi_2 = y_j) = p_{ij} \quad \cdot \quad \sum_{i,j} p_{ij} = 1$$

$$P(\xi_1 \leq x_i, \xi_2 \leq y_j) = \sum_{y \leq y_j} \sum_{x \leq x_i} p_{ij}$$

$$p_{i.} = P(\xi_1 = x_i) = \sum_j p_{ij}$$

$$p_{.j} = P(\xi_2 = y_j) = \sum_i p_{ij}$$

Asociando a la v.a. ξ_1 , un parámetro real cualquiera λ_1 , y a la v.a. ξ_2 , otro parámetro real λ_2 , tenemos por definición, una transvariación entre las v.a. ξ_1 y ξ_2 con respecto a los parámetros λ_1 y λ_2 , cuando en una realización determinada del e.a. Λ , en el cual se ha presentado el valor $z = (x, y)$ de la v.a. $\xi = (\xi_1, \xi_2)$ la diferencia $x - y$ es de signo opuesto al signo de la diferencia $\lambda_1 - \lambda_2$. El valor absoluto de la potencia $(x - y)^r$ nos define su respectiva intensidad de orden r , siendo $|x - y|$ la intensidad de primer orden y también la medida de dicha transvariación.

De entre todas las diferencias posibles $x - y$ que puedan presentarse de signo opuesto al signo de la diferencia $\lambda_1 - \lambda_2$, en la realización del e.a. Λ , la mayor en valor absoluto nos define

el campo de transvariación de las v.a. (ξ_1, ξ_2) con respecto a los parámetros (λ_1, λ_2) .

Para los casos en que sea $x - y = 0$, siendo $\lambda_1 \neq \lambda_2$, como así también para cuando sea $\lambda_1 = \lambda_2$, adoptamos una convención semejante a la introducida en 2.I.

3.2. — TRANSVARIABILIDAD

Definimos la transvariabilidad entre la v.a. bidimensional $\xi = (\xi_1, \xi_2)$ y el parámetro $\lambda = (\lambda_1, \lambda_2)$, a la probabilidad de que en una realización dada del e.a. Λ , la v.a. $\xi = (\xi_1, \xi_2)$ asuma un valor $z = (x, y)$ tal que el signo de la diferencia $x - y$ sea opuesto al signo de la diferencia $\lambda_1 - \lambda_2$.

De la definición de transvariabilidad se desprende que ella es igual a la probabilidad de que en una realización del e.a. Λ , la v.a. ξ_1 , presente un valor x menor que el valor y , correspondiente a la v.a. ξ_2 , si es $\lambda_1 > \lambda_2$ y a la inversa, si es $\lambda_1 < \lambda_2$. Es también la probabilidad de que la v.a. $\xi = (\xi_1, \xi_2)$ presente un valor $z = (x, y)$ dado por un punto en el semiplano de la izquierda (derecha) de la recta de ecuación $x - y = 0$, si el punto $\lambda = (\lambda_1, \lambda_2)$ se encuentra en el semiplano de la derecha (izquierda).

Simbólicamente tenemos (7)

$$p = \begin{cases} P(\xi_1 < \xi_2) = \int_{-\infty}^{\infty} dy \int_{-\infty}^y f(x, y) dx = \int_{-\infty}^{\infty} dx \int_x^{\infty} f(x, y) dy & \text{si } \lambda_1 > \lambda_2 \\ P(\xi_1 > \xi_2) = \int_{-\infty}^{\infty} dx \int_{-\infty}^x f(x, y) dy = 1 - P(\xi_1 < \xi_2) & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.1)$$

Para cada distribución de probabilidad, el valor de p es independiente de los valores de λ_1 y λ_2 , no así del signo de su diferencia, manteniéndose constante para cualquier valor de $\lambda = (\lambda_1, \lambda_2)$, perteneciente al mismo semiplano, de los dos que determina la recta de ecuación $x - y = 0$.

(7) Salvo especificación en contrario, se entiende, en todo lo que sigue, que trabajamos con v.a. que tienen una distribución continua de probabilidad.

Salvo el caso en que dicha recta sea una recta mediana de la v.a. $\xi = (\xi_1, \xi_2)$, es $p \neq \frac{1}{2}$ ⁽⁸⁾.

3.3. MAXIMO DE LA TRANSVARIABILIDAD

Definimos a continuación la transvariabilidad como función del parámetro real t , es decir, la probabilidad de que en la realización del e.a. Λ , la v.a. $\xi = (\xi_1, \xi_2)$ asuma un valor $z = (x, y)$ tal que el signo de la diferencia $x - (y + t)$ sea opuesto al signo de la diferencia $\lambda_1 - \lambda_2$. En símbolos

$$p(t) = \begin{cases} P(\xi_1 < \xi_2 + t) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{y+t} f(x, y) dx & \text{si } \lambda_1 > \lambda_2 \\ P(\xi_1 > \xi_2 + t) = 1 - P(\xi_1 < \xi_2 + t) & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.2)$$

siendo $P(\xi_1 < \xi_2 + t)$ una función de conjunto aditiva y no negativa del parámetro real t , por evidente consecuencia de las propiedades de toda función de distribución.

Nos interesa conocer el significado de t y determinar el máximo de $p(t)$ para cada una de las dos hipótesis precedentes. Para ello introduzcamos una recta de ecuación

$$(x - h) - (y - k) = 0 \quad (3.3)$$

con h y k reales y arbitrarios. Ella es paralela a la recta de ecuación

$$x - y = 0 \quad (3.4)$$

por tener sus respectivos coeficientes angulares iguales en valor y signo.

Si $p(t)$ es la probabilidad de que la v.a. $\xi = (\xi_1, \xi_2)$ asuma un valor $z = (x, y)$ perteneciente al semiplano determinado por (3.3) que no contenga el punto $\lambda = (\lambda_1, \lambda_2)$, es entonces $t = h - k$.

Simbolizando con R_2 el conjunto de valores que puede asumir la v.a. bidimensional ξ en todo el plano real y con S el sub-

⁽⁸⁾ DE LUCIA, LUIGI, *Rette Mediane*, «Rivista Italiana di Economia, Demografia e Statistica», Roma, N° 1, año 1953.

conjunto de R_2 perteneciente al semiplano determinado por (3.3) que no contenga el punto λ , todo punto del plano perteneciente a S nos define una transvariación, por ser siempre en él la diferencia $x - (y + t)$ de signo opuesto a $\lambda_1 - \lambda_2$. Su probabilidad es una función de conjunto aditiva y no negativa de S , cuyo máximo es alcanzado cuando la (3.3) pasa por el punto λ , siendo en tal caso $t = h - k = \lambda_1 - \lambda_2$, con lo que tenemos para la (3.3).

$$(x - \lambda_1) - (y - \lambda_2) = 0 \quad (3.5)$$

Cuando es $\lambda_1 > \lambda_2$, el punto λ se encuentra a la derecha de (3.4) y por lo tanto también de (3.3). En tal caso el campo de variabilidad de t es

$$-\infty < t < \lambda_1 - \lambda_2$$

y si es $\lambda_1 < \lambda_2$, el punto λ se encuentra a la izquierda de (3.4) y por lo tanto de (3.3). El campo de variabilidad de t es ahora

$$\lambda_1 - \lambda_2 < t < \infty$$

En consecuencia, simbolizando con p_M el máximo de la transvariabilidad, tenemos

$$p_M = \lim_{t \rightarrow \lambda_1 - \lambda_2} p(t) = \begin{cases} P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) = \int_{-\infty}^{\infty} dy \int_{-\infty}^{y + \lambda_1 - \lambda_2} f(x, y) dx & \text{si } \lambda_1 > \lambda_2 \\ P(\xi_1 > \xi_2 + \lambda_1 - \lambda_2) = 1 - P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.6)$$

siendo $p_M \neq \frac{1}{2}$ salvo el caso que la (3.5) sea una recta mediana de la v.a. ξ .

Al mismo resultado dado en (3.6) se llega analizando el comportamiento de la derivada de (3.2), en el intervalo de variabilidad de t . En efecto, para $\lambda_1 > \lambda_2$, se observa que $\frac{d p(t)}{d t}$ es una función no negativa de t y por lo tanto es $p(t)$ una función no decreciente, mientras que para $\lambda_1 < \lambda_2$, es $\frac{d p(t)}{d t}$ una función no positiva y en consecuencia es $p(t)$ una función no creciente de t ,

dentro de sus respectivos intervalos de variabilidad, cuyo máximo, en cada caso, está dado por (3.6).

El máximo de la transvariabilidad, como se desprende de la (3.6), es función del valor absoluto de la diferencia $\lambda_1 - \lambda_2$. Su campo de variabilidad está dado por

$$p \leq p_M \leq 1$$

siendo la transvariabilidad el menor valor que puede asumir p_M , que se varifica cuando es $\lambda_1 = \lambda_2$, y es 1 el máximo al que tiende p_M cuando $|\lambda_1 - \lambda_2| \rightarrow \infty$. Es suficiente para que sea $p_M = 1$, que λ_1 , sea mayor o igual que el mayor valor que pueda asumir la v.a. ξ_1 y que λ_2 sea menor o igual que el menor valor que pueda asumir la v.a. ξ_2 , cuando es $\lambda_1 > \lambda_2$. Si es $\lambda_1 < \lambda_2$, se cambian respectivamente las exigencias de λ_1 y λ_2 .

3.4. PROBABILIDAD DE TRANSVARIACION

Definimos la probabilidad de transvariación mediante el cociente entre la transvariabilidad y el máximo que ella pueda asumir. La necesidad de introducir este nuevo concepto ya ha sido dada en el capítulo precedente.

En nuestro caso, la probabilidad de transvariación entre la v.a. $\xi = (\xi_1, \xi_2)$ y los parámetros λ_1 y λ_2 es

$$P = \frac{p}{p_M} = \begin{cases} \frac{P(\xi_1 < \xi_2)}{P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2)} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^y f(x, y) dx}{\int_{-\infty}^{\infty} \int_{-\infty}^{y + \lambda_1 - \lambda_2} f(x, y) dx} & \text{si } \lambda_1 > \lambda_2 \\ \frac{P(\xi_1 > \xi_2)}{P(\xi_1 > \xi_2 + \lambda_1 - \lambda_2)} = \frac{1 - P(\xi_1 < \xi_2)}{1 - P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2)} & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.7)$$

Como características de la probabilidad de transvariación, en relación con la primera de (3.7), señalamos:

- 1) es función monótona no creciente de la diferencia $\lambda_1 - \lambda_2$;
- 2) es nula cuando el menor valor que puede asumir ξ_1 , en una realización cualquiera del e.a. Λ , es mayor que el mayor valor que puede asumir ξ_2 , siempre que sea $P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) \neq 0$;

3) Si se verifica también que $P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) = 0$ convenimos en definir igualmente como nula la probabilidad de transvariación ;

4) es igual a la transvariabilidad cuando λ_1 , es mayor que el mayor valor que puede asumir ξ_1 y λ_2 , es menor que el menor valor que puede asumir ξ_2 ;

5) es igual al doble de la transvariabilidad, cuando la (3.5) es una recta mediana de la v.a. $\xi = (\xi_1, \xi_2)$;

6) es igual a la unidad, que es el máximo valor que la probabilidad de transvariación puede asumir, cuando es $\lambda_1 = \lambda_2$ y $P(\xi_1 < \xi_2) \neq 0$;

7) convenimos en definir como nula la probabilidad de transvariación, cuando para $\lambda_1 = \lambda_2$, sea $P(\xi_1 < \xi_2) = 0$. Esta situación, si bien teóricamente concebible, prácticamente es de presentación imposible, con sólo recordar que, desde este punto de vista, trabajamos con λ_1 y λ_2 iguales a los respectivos valores medianos de las v.a. ξ_1 y ξ_2 , que es la hipótesis de trabajo más conveniente, si queremos extraer la mayor utilidad posible de nuestra investigación. La convención aquí adoptada modifica, a este solo efecto, lo convenido en 3.1 ;

8) es igual a la unidad cuando el mayor valor que puede asumir la v.a. ξ_1 , es menor que el menor valor que puede asumir la v.a. ξ_2 , es decir, cuando es $P(\xi_1 < \xi_2) = 1$;

9) es un valor abstracto, de dimensión cero, independiente de la unidad de medida en que vienen dadas las observaciones, por lo tanto apto para ser comparado con la probabilidad de transvariación de cualquier otro conjunto de observaciones, cualquiera sea su unidad de medida ;

10) su campo de variabilidad es $0 \leq P \leq 1$.

Modificando convenientemente las presentes características, obtenemos las correspondientes al supuesto $\lambda_1 < \lambda_2$.

3.5. MOMENTO DE TRANSVARIACIÓN

La suma de los valores absolutos de las intensidades de orden r de todas las transvariaciones posibles entre las v.a. ξ_1 y ξ_2 con respecto a los parámetros λ_1 y λ_2 , nos define el momento de transvariación de orden r .

En símbolos, tenemos

$$m_r = \begin{cases} \int_{-\infty}^{\infty} dy \int_{-\infty}^y (y-x)^r f(x, y) dx & \text{si } \lambda_1 > \lambda_2 \\ \int_{-\infty}^{\infty} dy \int_y^{\infty} (x-y)^r f(x, y) dx & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.8)$$

siendo el momento de orden cero igual a la transvariabilidad, es decir $m_0 = \phi$.

En cuanto al comportamiento de m_r , como función de r , no podemos anticipar nada y sólo deducirlo para cada distribución.

En efecto, derivando la (3.8), tenemos

$$\frac{d m_r}{d r} = \begin{cases} \int_{-\infty}^{\infty} dy \int_{-\infty}^y (y-x)^r \log (y-x) f(x, y) dx = \\ = \int_{-\infty}^{\infty} dy \int_{-\infty}^{y-1} (y-x)^r \log (y-x) f(x, y) dx + \\ + \int_{-\infty}^{\infty} dy \int_{y-1}^y (y-x)^r \log (y-x) f(x, y) dx & \text{si } \lambda_1 > \lambda_2 \\ \int_{-\infty}^{\infty} dy \int_y^{\infty} (x-y)^r \log (x-y) f(x, y) dx = \\ = \int_{-\infty}^{\infty} dy \int_{y+1}^{\infty} (x-y)^r \log (x-y) f(x, y) dx + \\ + \int_{-\infty}^{\infty} dy \int_y^{y+1} (x-y)^r \log (x-y) f(x, y) dx & \text{si } \lambda_1 < \lambda_2 \end{cases}$$

donde el último término de cada una de las dos igualdades precedentes es no positivo, mientras que el primero es no negativo, por lo tanto el signo de la suma depende, en cada una de las dos igualdades, del signo del sumando de mayor valor absoluto.

Analizando su derivada segunda

$$\frac{d^2 m_r}{d r^2} = \begin{cases} \int_{-\infty}^{\infty} d y \int_{-\infty}^y (y-x)^r \log^2 (y-x) f(x, y) d x \geq 0 & \text{si } \lambda_1 > \lambda_2 \\ \int_{-\infty}^{\infty} d y \int_y^{\infty} (x-y)^r \log^2 (x-y) f(x, y) d x \geq 0 & \text{si } \lambda_1 < \lambda_2 \end{cases}$$

que al darnos un valor no negativo para $\lambda_1 \gtrless \lambda_2$, concluimos que m_r puede ser una función constante, creciente, decreciente o presentar un mínimo.

El momento de transvariación como ya vimos para la probabilidad de transvariación, sólo depende, para cada distribución en particular, del signo de la diferencia existente entre λ_1 y λ_2 . Su valor no altera, cualquiera sea la ubicación de λ dentro de uno de los dos semiplanos determinados por la recta de ecuación $x - y = 0$

3.6. MAXIMO DEL MOMENTO DE TRANSVARIACIÓN

Definimos a continuación el momento de transvariación de orden r , en función del parámetro real t . Simbolizándolo con $m_r(t)$, es igual a la suma de los valores absolutos de la potencia r -ésima de $y - x + t$, de todos los casos posibles, en la realización del e.a. Λ , para los cuales la v.a. $\xi = (\xi_1, \xi_2)$ asuma un valor $z = (x, y)$ tal que el signo de la diferencia $x - (y + t)$ sea opuesto al signo de la diferencia $\lambda_1 - \lambda_2$. O sea

$$m_r(t) = \begin{cases} \int_{-\infty}^{\infty} d y \int_{-\infty}^{y+t} (y-x+t)^r f(x, y) d x & \text{si } \lambda_1 > \lambda_2 \\ \int_{-\infty}^{\infty} d y \int_{y+t}^{\infty} (x-y-t)^r f(x, y) d x & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.9)$$

donde, por lo ya visto en 3.3.

$$\begin{cases} -\infty < t < \lambda_1 - \lambda_2 & \text{si } \lambda_1 > \lambda_2 \\ \lambda_1 - \lambda_2 < t < \infty & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.10)$$

La función $m_r(t)$ es no decreciente para $\lambda_1 > \lambda_2$ y no creciente para $\lambda_1 < \lambda_2$, como se desprende de su derivada o simplemente del análisis de $m_r(t)$, variando t dentro de sus respectivos intervalos dados en (3.10).

Conduciendo el análisis en forma semejante al seguido en 3.3 y simbolizando con $m_{r,M}$ el máximo del momento de transvariación de orden r , tenemos

$$m_{r,M} = \begin{cases} \int_{-\infty}^{\infty} dy \int_{-\infty}^{y+\lambda_1-\lambda_2} [(y-\lambda_2) - (x-\lambda_1)]^r f(x, y) dx & \text{si } \lambda_1 > \lambda_2 \\ \int_{-\infty}^{\infty} dy \int_{y+\lambda_1-\lambda_2}^{\infty} [(x-\lambda_1) - (y-\lambda_2)]^r f(x, y) dx & \text{si } \lambda_1 < \lambda_2 \end{cases} \quad (3.11)$$

La (3.11) al igual que la (3.6) es función no decreciente del valor absoluto de la diferencia $\lambda_1 - \lambda_2$. Su campo de variabilidad oscila en el intervalo

$$m_r \leq m_{r,M} \leq E[|(\xi_1 - \lambda_1) - (\xi_2 - \lambda_2)|]^r$$

El valor mínimo m_r es alcanzado por $m_{r,M}$, cuando la diferencia $\lambda_1 - \lambda_2$ tiende a cero y el valor máximo, cuando λ_1 , es mayor que el mayor valor que pueda asumir la v.a. ξ_1 y λ_2 es menor que el menor valor que pueda asumir ξ_2 , si es $\lambda_1 > \lambda_2$ y a la inversa si es $\lambda_1 < \lambda_2$.

Simbolizando con $\beta_{r;\lambda_1,\lambda_2}$ el momento de orden r del valor absoluto de la diferencia $(\xi_1 - \lambda_1) - (\xi_2 - \lambda_2)$, tenemos, por definición de momentos

$$\beta_{r;\lambda_1,\lambda_2} = E[|(\xi_1 - \lambda_1) - (\xi_2 - \lambda_2)|]^r \quad (3.12)$$

Cuando λ_1 y λ_2 son iguales, respectivamente, a las medias aritméticas $\alpha_{1.0}$ y $\alpha_{0.1}$ de las v.a. ξ_1 y ξ_2 , tenemos

$$\beta_{r;\alpha_{1.0},\alpha_{0.1}} = E(|\delta_1 - \delta_2|)^r \quad (3.13)$$

siendo δ_1 y δ_2 sus correspondientes v.a. desvíos.

Para $\lambda_1 = \lambda_2 = 0$, tenemos

$$\beta_{r,0,0} = E(|\xi_2 - \xi_1|)^r \quad (3.14)$$

Para $r = 1$, la (3.13) nos define la diferencia media de Gini.

3.7. INTENSIDAD DE TRANSVARIACIÓN

Definimos la intensidad de transvariación de orden r , mediante la raíz r -ésima del cociente entre el momento de transvariación de orden r y el máximo que el mismo puede alcanzar. Simbolizándola con I_r , tenemos

$$I_r = \frac{\frac{1}{m_{r,M}^r}}{\frac{1}{m_{r,M}^r}} = \frac{\left(\int_{-\infty}^{\infty} dy \int_{-\infty}^y (y-x)^r f(x,y) dx \right)^{\frac{1}{r}}}{\left(\int_{-\infty}^{\infty} dy \int_{y+\lambda_1-\lambda_2}^{\infty} [(y-\lambda_2) - (x-\lambda_1)]^r f(x,y) dx \right)^{\frac{1}{r}}} \text{ si } \lambda_1 > \lambda_2 \quad (3.15)$$

$$I_r = \frac{\frac{1}{m_{r,M}^r}}{\frac{1}{m_{r,M}^r}} = \frac{\left(\int_{-\infty}^{\infty} dy \int_y^{\infty} (x-y)^r f(x,y) dx \right)^{\frac{1}{r}}}{\left(\int_{-\infty}^{\infty} dy \int_{y+\lambda_1-\lambda_2}^{\infty} [(x-\lambda_1) - (y-\lambda_2)]^r f(x,y) dx \right)^{\frac{1}{r}}} \text{ si } \lambda_1 < \lambda_2$$

Como características de la intensidad de transvariación, considerando el caso $\lambda_1 > \lambda_2$, tenemos

1) es función no creciente del valor absoluto de la diferencia $\lambda_1 - \lambda_2$;

2) es nula cuando el menor valor que puede asumir ξ_1 , en una realización del e.a. Λ , es mayor que el mayor valor que puede asumir ξ_2 , es decir, cuando es $P(\xi_1 < \xi_2) = 0$, siempre que sea $P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) \neq 0$;

3) si es, simultáneamente, $P(\xi_1 < \xi_2) = 0$ y $P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) = 0$, convenimos en definir como nula la intensidad de transvariación;

4) es

$$I_r = \frac{m_r^{\frac{1}{r}}}{\beta_{r|\lambda_1, \lambda_2}} = \frac{m_{(r)}}{\{E[|(\xi_1 - \lambda_1) - (\xi_2 - \lambda_2)|]\}^{\frac{1}{r}}}$$

cuando λ_1 y λ_2 son valores tales que en la realización del e.a. Λ , ningún valor de ξ_1 , supera a λ_1 , y los de ξ_2 son todos superiores a λ_2 ;

5) para λ_1 y λ_2 iguales a las medias aritméticas, respectivamente, de las v.a. ξ_1 y ξ_2 y si estas v.a. obedecen a una ley simétrica de distribución, es

$$I_r = \frac{m_{(r)}}{m_{(r;M)}} = \frac{(2 m_r)^{\frac{1}{r}}}{\beta_{r, a_{10}, a_{01}}^{\frac{1}{r}}}$$

6) es igual a la unidad que es el máximo valor que la intensidad de transvariación puede asumir, cuando es $\lambda_1 = \lambda_2$ y $P(\xi_1 < \xi_2) \neq 0$;

7) convenimos en definir como nula la intensidad de transvariación cuando para $\lambda_1 = \lambda_2$, sea $P(\xi_1 < \xi_2) = 0$, con lo que modificamos a este solo efecto, dada la indeterminación que se presenta, lo convenido en 3.1;

8) es

$$I_r = \frac{\beta_{r;0,0}^{\frac{1}{r}}}{\beta_{r;\lambda_1, \lambda_2}^{\frac{1}{r}}}$$

cuando es $P(\xi_1 < \xi_2) = 1$

9) es un valor de dimensión cero, independiente de la unidad de medida en que vienen dadas las observaciones, por lo tanto apto para ser comparado con la intensidad de transvariación del mismo orden de cualquier otro conjunto de observaciones, cualquiera sea su unidad de medida;

10) oscila en el intervalo $0 \leq I_r \leq 1$.

Modificando convenientemente las presentes características, obtenemos las correspondientes al supuesto $\lambda_1 < \lambda_2$. Sin pérdida

de generalidad y salvo expresa indicación en contrario, suponemos en adelante que $\lambda_1 > \lambda_2$.

3.8. REDUCCION A FORMULAS EN R_1 DE LOS PARAMETROS DE LA TRANSVARIACION EN R_2

La teoría de la transvariación no escapa, como no escapa ningún otro capítulo del Método Estadístico, al crecimiento de la complejidad cuando aumenta el número de dimensiones de la *v.a.* considerada. Si bien, desde el punto de vista teórico, esta situación se supera sin mayores inconvenientes, desde el punto de vista operacional, cuando encaramos la aplicación de la teoría a fenómenos reales, la complejidad del cálculo termina por volver prohibitivo el trabajo. Esta situación en nuestro caso, se presenta no solo cuando crece el número de dimensiones de la *v.a.* sino también, dentro de una dada dimensión, cuando pasamos de la probabilidad a la intensidad de transvariación. Esta situación ha sido abordada con éxito por Gini (Refs. 13 y 15), Livada (Ref. 19), De Lucía (Ref. 9) y Dahlberg.

Una sencilla transformación lineal nos conducirá, para los parámetros de la transvariación en R_2 , a fórmulas de fácil aplicación a cada distribución considerada y en especial simplificará el trabajo operatorio cuando nos ocupemos de la transvariación en el muestreo.

Introducimos una nueva *v.a.*, definida como sigue

$$\eta = \xi_2 - \xi_1 \quad (3.16)$$

siendo $\lambda_1 > \lambda_2$, hipótesis con la que, sin pérdida de generalidad, trabajamos en lo sucesivo, como ya señalamos al final de 3.7.

Simbolizando con t el valor que asume la *v.a.* $\eta = \xi_2 - \xi_1$, en una realización del e.a. Λ , y con $g(t)$ a su f. de d., es

$$t = y - x \quad (3.16.a)$$

$$g(t) = \int_{-\infty}^{\infty} f(y - t, y) dy \quad (3.17)$$

En nuestra hipótesis, o sea $\lambda_1 > \lambda_2$, todo valor positivo de t , nos define una transvariación, luego, la transvariabilidad es igual a la probabilidad de que en una realización del e.a. Λ , la v.a. $\eta = \xi_2 - \xi_1$, asuma un valor $t = y - x > 0$, es decir

$$p = P(\eta > 0) = \int_0^{\infty} g(t) dt \quad (3.18)$$

siendo el máximo de la transvariabilidad igual a

$$p_M = P(\eta > \lambda_2 - \lambda_1) = \int_{-(\lambda_1 - \lambda_2)}^{\infty} g(t) dt \quad (3.19)$$

El momento de transvariación de orden r de la v.a. η , es, por definición

$$m_r = \int_0^{\infty} t^r g(t) dt \quad (3.20)$$

y el máximo de dicho momento

$$m_{r;M} = \int_{-(\lambda_1 - \lambda_2)}^{\infty} (t + \lambda_1 - \lambda_2)^r g(t) dt \quad (3.21)$$

Por definición de probabilidad de transvariación obtenemos, a partir de (3.18) y (3.19)

$$P = \frac{p}{p_M} = \frac{P(\eta > 0)}{P(\eta > \lambda_2 - \lambda_1)} = \frac{\int_0^{\infty} g(t) dt}{\int_{-(\lambda_1 - \lambda_2)}^{\infty} g(t) dt} \quad (3.22)$$

y por definición de intensidad de transvariación de orden r , obtenemos a partir de (3.20) y (3.21)

$$I_r = \frac{m_{(r)}}{m_{(r;M)}} = \frac{\left(\int_0^\infty t^r g(t) dt \right)^{\frac{1}{r}}}{\left(\int_{(\lambda_1 - \lambda_2)}^\infty (t + \lambda_1 - \lambda_2)^r g(t) dt \right)^{\frac{1}{r}}} \quad (3.23)$$

3.9. TRANSVARIACION EN LA HIPOTESIS DE DISTRIBUCION NORMAL BIDIMENSIONAL

Dado el e.a. Λ , al cual asociamos la v.a. $\xi = (\xi_1, \xi_2)$, distribuída normalmente en R_2 , siendo R_2 todo el plano real, con media y varianza $\alpha_{1.0}$ y σ_1^2 , para ξ_1 y $\alpha_{0.1}$ y σ_2^2 para ξ_2 , su f. de d. es

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[\frac{(x-\alpha_{1.0})^2}{\sigma_1^2} - \frac{2\rho(x-\alpha_{1.0})(y-\alpha_{0.1})}{\sigma_1\sigma_2} + \frac{(y-\alpha_{0.1})^2}{\sigma_2^2} \right]} \quad (3.24)$$

a) *Primer método.* — Deduciremos a continuación los parámetros de la transvariación de las v.a. ξ_1 y ξ_2 , cuya f. de d. conjunta está dada por la (3.24). Supondremos, evidentemente sin pérdida de generalidad que $\alpha_{1.0} > \alpha_{0.1}$, haciendo

$$d = \alpha_{1.0} - \alpha_{0.1} > 0$$

Por definición de transvariabilidad, véase (3.1), es

$$p = P(\xi_1 < \xi_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^\infty dy \int_{-\infty}^y e^{-\frac{1}{2(1-\rho^2)} A} dx \quad (3.25)$$

y por definición de momento de transvariación de orden r , véase (3.8), es

$$m_r = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^\infty dy \int_{-\infty}^y (y-x)^r e^{-\frac{1}{2(1-\rho^2)} A} dx \quad (3.26)$$

siendo

$$A = \frac{(x - \alpha_{1.0})^2}{\sigma_1^2} - 2 \rho \frac{(x - \alpha_{1.0})(y - \alpha_{0.1})}{\sigma_1 \sigma_2} + \frac{(y - \alpha_{0.1})^2}{\sigma_2^2}$$

Haciendo la sustitución

$$\begin{cases} u = \frac{x - \alpha_{1.0}}{\sigma_1} \\ v = \frac{y - \alpha_{0.1}}{\sigma_2} \end{cases} \quad (3.27)$$

es decir, standardizando las *v.a.* ξ_1 y ξ_2 , y siendo el jacobiano de la sustitución

$$J = \sigma_1 \sigma_2$$

tenemos, respectivamente, para la (3.25) y (3.26)

$$p = \frac{1}{2 \pi \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} dv \int_{-\infty}^{\frac{\sigma_2}{\sigma_1} v - \frac{d}{\sigma_1}} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} du \quad (3.28)$$

$$m_r = \frac{1}{2 \pi \sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} dv \int_{-\infty}^{\frac{\sigma_2}{\sigma_1} v - \frac{d}{\sigma_1}} (\sigma_2 v - \sigma_1 u - d)^r e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} du \quad (3.29)$$

La (3.28) - la (3.29) recibe una interpretación análoga — representa geoméricamente la parte del volumen unidad comprendido entre el plano (u, v) y la f. de d.

$$z = f(u, v) = \frac{1}{2 \pi \sqrt{1 - \rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)} \quad (3.30)$$

extendido a la izquierda del plano secante, paralelo al eje Z , cuya intersección con el plano (u, v) , determina las rectas de ecuación

$$\begin{cases} u = \frac{\sigma_2}{\sigma_1} v - \frac{d}{\sigma_1} \\ v = \frac{\sigma_1}{\sigma_2} u + \frac{d}{\sigma_2} \end{cases}$$

Entendemos por parte izquierda del volumen unidad dividido por el plano secante la que corresponde a un observador ubicado en el origen y mirando el eje v en su dirección positiva.

Efectuando una rotación del sistema (u, v) en una medida igual al ángulo θ , las coordenadas del sistema (u, v) en función del nuevo sistema (u', v') , son

$$\begin{cases} u = u' \cos \cdot \theta - v' \sin \cdot \theta \\ v = u' \sin \cdot \theta + v' \cos \cdot \theta \end{cases} \quad \dots \quad \begin{cases} u' = u \cos \cdot \theta + v \sin \cdot \theta \\ v' = v \cos \cdot \theta - u \sin \cdot \theta \end{cases} \quad (3.31)$$

con

$$\text{tag } \theta = - \frac{\sigma_2}{\sigma_1} \quad (3.32)$$

resultando para el exponente de (3.28) y (3.29), en función de las coordenadas del nuevo sistema y teniendo presente (3.32)

$$\frac{1}{2(1 - \rho^2)} (u'^2 - 2\rho v'u + v^2) = \frac{1}{2(1 - \rho^2)} \cdot \left(u'^2 \frac{\sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} - 2\rho u'v' \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2} + v'^2 \frac{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} \right)$$

Siendo igual a uno el jacobiano de la sustitución, tenemos para la transvariabilidad y el momento de transvariación de orden r , dados respectivamente por (3.28) y (3.29)

$$p = \frac{1}{2\pi\sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} dv' \int_{-\infty}^{-h} e^{-(au'^2 + 2bu'v' + cv'^2)} dv' \quad (3.33)$$

$$m_r = \frac{1}{2\pi\sqrt{1 - \rho^2}} \int_{-\infty}^{\infty} dv' \int_{-\infty}^{-h} (-d - u'\sqrt{\sigma_1^2 + \sigma_2^2})^r e^{-(au'^2 + 2bu'v' + cv'^2)} dv' \quad (3.34)$$

siendo

$$\left\{ \begin{aligned} a &= \frac{\sigma_1^2 + \sigma_2^2 + 2 \rho \sigma_1 \sigma_2}{2 (1 - \rho^2) (\sigma_1^2 + \sigma_2^2)} \\ b &= - \frac{\rho (\sigma_1^2 - \sigma_2^2)}{2 (1 - \rho^2) (\sigma_1^2 + \sigma_2^2)} \\ c &= \frac{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}{2 (1 - \rho^2) (\sigma_1^2 + \sigma_2^2)} \\ h &= \frac{d}{\sqrt{\sigma_1^2 + \sigma_2^2}} \end{aligned} \right. \quad (3.35)$$

donde h es la distancia desde el origen a la recta de ecuación

$$u = \frac{\sigma_2}{\sigma_1} v - \frac{d}{\sigma_1}$$

Realizamos ahora una nueva rotación de ejes, sometida a las siguientes condiciones

1) con respecto al nuevo sistema de ejes, que los simbolizamos (x, y) , es nulo el coeficiente de correlación;

2) el intervalo de integración subsiste, como en (3.33) y (3.34), independiente de la variable de integración.

Introduciendo los parámetros incógnitas k, l, m y n , para los cuales es

$$\begin{cases} u' = k x + l y \\ v' = m x + n y \end{cases} \quad (3.36)$$

La condición 1) se expresa mediante la siguiente identidad

$$\frac{1}{2} (x^2 + y^2) = a u'^2 + 2 b u' v' + c v'^2 \quad (3.37)$$

mientras que por la condición 2) es

$$l = 0$$

con lo que obtenemos para k , m y n , a partir de (3.37) y teniendo presente la (3.35) y (3.36)

$$\left\{ \begin{aligned} k &= -\sqrt{\frac{c}{2(a c - b^2)}} = -\sqrt{\frac{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}{\sigma_1^2 + \sigma_2^2}} \\ m &= \frac{b}{\sqrt{2 c (a c - b^2)}} = -\frac{\rho (\sigma_1^2 - \sigma_2^2)}{\sqrt{(\sigma_1^2 + \sigma_2^2) (\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2)}} \\ n &= \frac{1}{\sqrt{2 c}} = \frac{\sqrt{(1 - \rho^2) (\sigma_1^2 + \sigma_2^2)}}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}} \end{aligned} \right. \quad (3.38)$$

siendo

$$J = -\sqrt{1 - \rho^2}$$

el jacobiano de la transformación (3.36), tenemos para la (3.33) y (3.34), respectivamente

$$p = \frac{1}{2 \pi} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{d_1}^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{\sqrt{2 \pi}} \int_{d_1}^{\infty} e^{-\frac{x^2}{2}} dx \quad (3.39)$$

$$m_r = \frac{1}{2 \pi} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy \int_{d_1}^{\infty} (-d + x \sqrt{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}) e^{-\frac{x^2}{2}} dx = \quad (3.40)$$

$$= \frac{1}{\sqrt{2 \pi}} \int_{d_1}^{\infty} (-d + x \sqrt{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}) e^{-\frac{x^2}{2}} dx$$

siendo

$$d_1 = \frac{d}{\sqrt{\sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2}} \quad (3.41)$$

Para la transvariabilidad, dada por la (3.39), que no es otra cosa que el momento de transvariación de orden cero, tenemos en definitiva

$$p = 1 - \Phi(d_1) = \Phi(-d_1) \quad (3.42)$$

donde $\Phi(x)$ en la f. de p.a. de la v.a. standardizada que definimos en (2.38)

A partir de la (3.40), tenemos

$$\begin{aligned}
 m_r &= (-d)^r \Phi(-d_1) + \\
 &+ \frac{1}{\sqrt{2\pi}} \sum_{s=1}^r \binom{r}{s} (-d)^{r-s} (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{\frac{s}{2}} \int_{d_1}^{\infty} x^s e^{-\frac{x^2}{2}} dx = \quad (3.43) \\
 &= (-d)^r \Phi(-d_1) + \frac{1}{\sqrt{2\pi}} \sum_{s=1}^r \binom{r}{s} (-d)^{r-s} (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{\frac{s}{2}} G(s; d_1)
 \end{aligned}$$

donde $G(s; d_1)$, para s par e impar, lo tenemos resuelto, respectivamente, en (2.48) y (2.47).

El máximo de la transvariabilidad y del momento de transvariación de orden r , lo obtenemos, como ya vimos, mediante la sustitución

$$\begin{cases} u = x - \lambda_1 \\ v = y - \lambda_2 \end{cases}$$

Habiendo dado a λ_1 y λ_2 , en este párrafo, los valores correspondientes a las medias, respectivamente, de las v.a. ξ_1 y ξ_2 , la sustitución será

$$\begin{cases} u = x - \alpha_{1.0} \\ v = y - \alpha_{0.1} \end{cases}$$

lo que equivale a trabajar con sus correspondientes v.a. desvíos. El valor d es nulo para el máximo de la transvariabilidad y el máximo del momento de transvariación de orden r , con lo que tenemos para ellos, a partir de (3.39) y (3.40)

$$p_M = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = \frac{1}{2} \quad (3.44)$$

$$\begin{aligned}
 m_{r;M} &= \frac{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{\frac{r}{2}}}{\sqrt{2\pi}} \int_0^\infty x^r e^{-\frac{x^2}{2}} dx = \\
 &= \frac{(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{\frac{r}{2}}}{2\sigma^r} \beta_r = \frac{2^{\frac{r-2}{2}} (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{\frac{r}{2}}}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right)
 \end{aligned} \tag{3.45}$$

donde β_r , nos define, como vimos en (2.41), el momento absoluto de orden r de la v.a. normal $(0, \sigma)$. Su resultado lo tomamos de (2.50) y (2.51). El valor de β_r , para r par e impar, está dado por (2.51.a) y (2.51.b), respectivamente.

A partir de (3.42) y (3.44), tenemos, por definición, para la probabilidad de transvariación

$$P = \frac{\dot{p}}{\dot{p}_M} = 2 \Phi(-d_1) \tag{3.46}$$

A partir de la (3.43) y (3.45), por definición, tenemos para la intensidad de transvariación de orden r

$$\begin{aligned}
 I_r = & \frac{\left[(-d)^r \Phi(-d_1) + \frac{1}{\sqrt{2\pi}} \sum_{s=1}^r \binom{r}{s} (-d)^{r-s} (\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)^{\frac{s}{2}} G(s; d_1) \right]^{\frac{1}{r}}}{\sqrt{2(\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2)} \left[\frac{1}{2\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \right]^{\frac{1}{r}}}
 \end{aligned} \tag{3.47}$$

Para la intensidad de transvariación de orden 1, tenemos

$$I_1 = \frac{m_1}{m_{1;M}} = \sqrt{2\pi} [\Phi'(d_1) - d_1 \Phi(-d_1)] = \sqrt{2\pi} [\Phi'(d_1) - d_1 \dot{p}] \tag{3.48}$$

b) *Segundo metodo.* — Evidentemente en forma más inmediata arribamos al resultado de la probabilidad e intensidad de transvariación entre dos v.a. normalmente correlacionadas, mediante la transformación lineal vista en el párrafo 3.8.

Siendo ξ_1 y ξ_2 , dos v.a. normalmente correlacionadas, con medias $\alpha_{1.0}$ y $\alpha_{0.1}$ y varianzas σ_1 y σ_2 , la v.a. η , dada por (3.16),

que es función lineal de *v.a.* normales, es una nueva *v.a.* normalmente distribuida, siendo su correspondiente media y varianza

$$\begin{cases} \alpha = \alpha_{0.1} - \alpha_{1.0} = -d < 0 \\ \sigma^2 = \sigma_1^2 + \sigma_2^2 - 2 \rho \sigma_1 \sigma_2 \end{cases} \quad (3.49)$$

Su f. de d. es

$$g(t) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(t-\alpha)^2}{2\sigma^2}} \quad (3.50)$$

Reemplazando estos valores en (3.18), (3.19), (3.20) y (3.21), teniendo presente que trabajamos con $\lambda_1 = \alpha_{1.0}$ y $\lambda_2 = \alpha_{0.1}$, obtenemos respectivamente para la transvariabilidad, máximo de la transvariabilidad, momento de transvariación de orden *r* y su máximo, luego de efectuar la sustitución

$$u = \frac{t - \alpha}{\sigma}$$

$$p = P(\eta > 0) = \frac{1}{\sqrt{2\pi}} \int_{-\frac{\alpha}{\sigma}}^{\infty} e^{-\frac{u^2}{2}} du = \Phi\left(\frac{\alpha}{\sigma}\right) \quad (3.51)$$

$$p_M = P(\eta > \alpha) = \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{u^2}{2}} du = \frac{1}{2} \quad (3.52)$$

$$\begin{aligned} m_r &= \frac{1}{\sqrt{2\pi}} \int_{-\frac{\alpha}{\sigma}}^{\infty} (\alpha + \sigma u)^r e^{-\frac{u^2}{2}} du = \frac{1}{\sqrt{2\pi}} \sum_{s=0}^r \binom{r}{s} \alpha^{r-s} \sigma^s \int_{-\frac{\alpha}{\sigma}}^{\infty} u^s e^{-\frac{u^2}{2}} du = \\ &= \alpha^r \Phi\left(\frac{\alpha}{\sigma}\right) + \frac{1}{\sqrt{2\pi}} \sum_{s=1}^r \binom{r}{s} \alpha^{r-s} \sigma^s G\left(s; -\frac{\alpha}{\sigma}\right) \end{aligned} \quad (3.53)$$

donde el valor de $G\left(s; -\frac{\alpha}{\sigma}\right)$, para s par e impar, está dado por (2.48) y (2.47) respectivamente.

$$m_{r;M} = \frac{\sigma^r}{\sqrt{2\pi}} \int_0^\infty u^r e^{-\frac{u^2}{2}} du = \frac{\beta_r}{2} = \frac{2^{\frac{r-2}{2}} \sigma^r}{\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \quad (3.54)$$

donde el valor de la integral la obtenemos a partir de (2.50), el de β_r , a partir de (2.51) y el de β_r para r par o impar a partir de (2.51.a) y (2.51.b), respectivamente.

Por definición de probabilidad de transvariación, tenemos, a partir de (3.51) y (3.52)

$$P = \frac{\dot{p}}{\dot{p}_M} = 2 \Phi\left(\frac{\alpha}{\sigma}\right) \quad (3.55)$$

y teniendo presente (3.41) y (3.49) es

$$P = 2 \Phi(-d_1) \quad (3.56)$$

que coincide con (3.46)

Por definición de intensidad de transvariación de orden r , tenemos, a partir de (3.53) y (3.54)

$$I_r = \frac{\frac{1}{m_{r;M}^r}}{\sigma \sqrt{2} \left[\frac{1}{2\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \right]^{\frac{1}{r}}} = \frac{\left[\alpha^r \Phi\left(\frac{\alpha}{\sigma}\right) + \frac{1}{\sqrt{2\pi}} \sum_{s=1}^r \binom{r}{s} \alpha^{r-s} \sigma^s G\left(s; -\frac{\alpha}{\sigma}\right) \right]^{\frac{1}{r}}}{\sigma \sqrt{2} \left[\frac{1}{2\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \right]^{\frac{1}{r}}} \quad (3.57)$$

Teniendo presente los valores de α y σ dados por (3.49) y el valor de d_1 , dado por (3.41), llegamos al resultado (3.47), obtenido por el primer método.

Para $r=1$ tenemos evidentemente el resultado (3.48), que con la notación (3.49) es:

$$\begin{aligned} I_1 &= \frac{m_1}{m_{1;M}} = \sqrt{2\pi} \left[\Phi'\left(\frac{\alpha}{\sigma}\right) + \frac{\alpha}{\sigma} \Phi\left(\frac{\alpha}{\sigma}\right) \right] = \\ &= \sqrt{2\pi} \left[\Phi'\left(\frac{\alpha}{\sigma}\right) + \frac{\alpha}{\sigma} \dot{p} \right] \end{aligned} \quad (3.58)$$

Recordamos el carácter de función par de la f. de d. de la v.a. normal $(0, \sigma)$ y normal $(0, 1)$. Luego

$$\Phi' \left(\frac{\alpha}{\sigma} \right) = \Phi' \left(-\frac{\alpha}{\sigma} \right) = \Phi' (d_1) = \Phi' (-d_1)$$

Las características de la probabilidad de transvariación y de la intensidad de transvariación, se deducen fácilmente de las características dadas para el caso general, en 3.4 y 3.7, teniendo presente la simetría de la distribución normal bidimensional definida en todo el plano real, y la circunstancia de haber trabajado con los valores $\alpha_{1.0}$ y $\alpha_{0.1}$ para los parámetros λ_1 y λ_2 , respectivamente. Por ser la recta de ecuación $(x - \alpha_{1.0}) - (y - \alpha_{0.1}) = 0$, una recta mediana de la distribución, hemos obtenido el valor un medio para el máximo de la transvariabilidad.

4. — TRANSVARIACIÓN ENTRE DOS VARIABLES ALEATORIAS INDEPENDIENTES

4.1. INTRODUCCIÓN

Un caso particular de la transvariación en R_2 lo constituye la hipótesis de independencia entre las v.a. ξ_1 y ξ_2 , componentes de la v.a. bidimensional ξ .

La condición necesaria y suficiente para que las v.a. ξ_1 y ξ_2 , componentes de la v.a. ξ , adscripta al e.a. Λ , sean independientes, es que la f. de p.a. de la v.a. bidimensional sea igual al producto de las f. de p.a. de las v.a. que la componen; o sea

$$F(x, y) = P(\xi_1 \leq x, \xi_2 \leq y) = F_1(x) F_2(y) \quad (4.1)$$

En función de sus respectivas f. de d., para los casos de dos v.a. de tipo continuo y de tipo discreto, respectivamente tenemos

$$\left\{ \begin{array}{l} f(x, y) = P(\xi_1 = x, \xi_2 = y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = f_1(x) f_2(y) \\ P(\xi_1 = x_i, \xi_2 = y_j) = p_{1i} p_{2j} \end{array} \right. \quad (4.1.a)$$

donde los factores de (4.1) y (4.1.a) nos definen, respectivamente, las f. de p.a. y las f. de d. de las distribuciones marginales de la v.a. bidimensional.

La f. de p.a. de la suma de las v.a. independientes ξ_1 y ξ_2 donde $t = x + y$ es una particular determinación de la v.a. $\eta = \xi_1 + \xi_2$, es

$$F(t) = E[F_2(\eta - \xi_1)] = E[F_1(\eta - \xi_2)] \quad (4.2)$$

Para el caso que por lo menos una de las v.a. componentes sea de tipo continuo, es

$$F(t) = \int_{-\infty}^{\infty} F_2(t - x) dF_1(x) = \int_{-\infty}^{\infty} F_1(t - y) dF_2(y) \quad (4.2.a)$$

Para el caso objeto de aplicación en el presente capítulo, es decir la f. de p.a. de la v.a.

$$\eta = \xi_2 - \xi_1$$

tenemos

$$F(t) = P(\xi_2 - \xi_1 \leq \eta) = \int_{-\infty}^{\infty} F_2(t + x) dF_1(x) \quad (4.3)$$

Si la (4.3) admite derivada en el punto t , tenemos para la f. de d. de la v.a. $\eta = \xi_2 - \xi_1$

$$f(t) = P(\xi_2 - \xi_1 = \eta) = \int_{-\infty}^{\infty} f_2(t + x) dF_1(x) \quad (4.3.a)$$

El momento natural de orden $r + s$ de la v.a. $\xi = (\xi_1, \xi_2)$, tomada desde el origen, es, por definición de momento producto y teniendo presente el carácter de independencia de las v.a. ξ_1, ξ_2

$$\alpha_{r,s} = E(\xi_1^r, \xi_2^s) = E(\xi_1^r) E(\xi_2^s) = \alpha_{r,0} \cdot \alpha_{0,s} \quad (4.4)$$

Los conceptos y definiciones dados en el capítulo precedente, sobre transvariación, su intensidad de orden r , su medida, transvariabilidad, máximo de la transvariabilidad, momento de transvariación de orden r , máximo de este momento, probabilidad de transvariación e intensidad de transvariación de orden r , subsisten plenamente en el presente capítulo, simplificándose solo su estructura analítica por la circunstancia de que aquí trabajamos con variables aleatorias independientes.

A los parámetros de la transvariación de los que nos venimos ocupando, agregamos los introducidos por Gini y Livada (Ref. 15) a saber: área de transvariación, relación de transvariación, valor discriminativo y valor crítico.

4.2. PROBABILIDAD DE TRANSVARIACIÓN

Teniendo presente la (4.1.a), tenemos para la transvariabilidad, su valor máximo y para la probabilidad de transvariación, a partir de (3.1), (3.6) y (3.7), respectivamente

$$p = P(\xi_1 < \xi_2) = \int_{-\infty}^{\infty} F_1(y) dF_2(y) \quad (4.5)$$

$$p_M = P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2) = \int_{-\infty}^{\infty} F_1(y + \lambda_1 - \lambda_2) dF_2(y) \quad (4.6)$$

$$P = \frac{p}{p_M} = \frac{P(\xi_1 < \xi_2)}{P(\xi_1 < \xi_2 + \lambda_1 - \lambda_2)} \quad (4.7)$$

Si λ_1 y λ_2 , son sus respectivos valores medianos, es $P = 2p$.

Si las v.a. ξ_1 y ξ_2 son v.a. idénticas, tenemos para la (4.5), por razonamiento directo o integrando por partes

$$p = \int_{-\infty}^{\infty} F_1(y) dF_1(y) = \frac{1}{2} \quad (4.8)$$

La transvariabilidad en esta hipótesis, nos da la probabilidad que en dos realizaciones independientes del e.a. Λ , la v.a. ξ_1 asuma en la segunda prueba un valor mayor que en la primera.

Introduciendo la sustitución (3.16), tenemos para la (3.17), por tratarse de v.a. independientes

$$g(t) = \int_{-\infty}^{\infty} f_1(x-t) dF_2(x) \quad (4.9)$$

a partir de la cual obtenemos, como ya vimos, la probabilidad de transvariación.

4.3. INTENSIDAD DE TRANSVARIACIÓN

A partir de (3.8), (3.11) y (3.15) y teniendo presente la (4.1.a), obtenemos para el nomento de transvariación de orden r , su valor máximo y para la intensidad de transvariación de orden r , respectivamente

$$m_r = \int_{-\infty}^{\infty} f_2(y) dy \int_{-\infty}^y (y-x)^r f_1(x) dx \quad (4.10)$$

$$m_{r,M} = \int_{-\infty}^{\infty} f_2(y) dy \int_{-\infty}^{y+\lambda_1-\lambda_2} [(y-\lambda_2) - (x-\lambda_1)]^r f_1(x) dx \quad (4.11)$$

$$I_r = \frac{(m_r)^{\frac{1}{r}}}{(m_{r,M})^{\frac{1}{r}}} = \frac{m_{(r)}}{m_{(r;M)}} \quad (4.12)$$

Con la sustitución (3.16) tenemos respectivamente, para los parámetros aquí considerados, la (3.20), (3.21) y (3.23), donde la f. de d. está ahora dada por (4.9).

Las características de la probabilidad e intensidad de transvariación, se obtienen a partir de las ya dadas en el capítulo precedente.

4.4. ÁREA DE TRANSVARIACIÓN

Sean N_1 y N_2 respectivamente el número de elementos componentes de los colectivos sobre los cuales se realiza el e.a. Λ , al

cual se encuentra adscripta la v.a. $\xi = (\xi_1, \xi_2)$, siendo ξ_1 y ξ_2 v.a. independientes con f. de d. $f_1(x)$ y $f_2(x)$ respectivamente.

Introduzcamos ahora la v.a. η , perteneciente al colectivo de tamaño $N = N_1 + N_2$, cuya f. de d. es

$$f(x) = \frac{N_1 f_1(x) + N_2 f_2(x)}{N_1 + N_2} \quad (4.13)$$

Asociamos a esta v.a. η , la función de frecuencia $g(x)$, la cual definimos de la siguiente manera

$$\begin{cases} g(x) = N_1 f_1(x) & \text{si es } N_1 f_1(x) \leq N_2 f_2(x) \\ g(x) = N_2 f_2(x) & \text{si es } N_1 f_1(x) > N_2 f_2(x) \end{cases} \quad (4.14)$$

La (4.13) nos da la probabilidad de que en una realización del e.a. Λ , la v.a. η asuma el valor x , mientras que, para este mismo valor de la v.a. η , es

$$\frac{g(x)}{N_1 + N_2} = \frac{g(x)}{N} \quad (4.15)$$

la probabilidad de que x se a una particular determinación de la v.a. ξ_1 o de la ξ_2 , según sea $N_1 f_1(x)$ no mayor o mayor de $N_2 f_2(x)$.

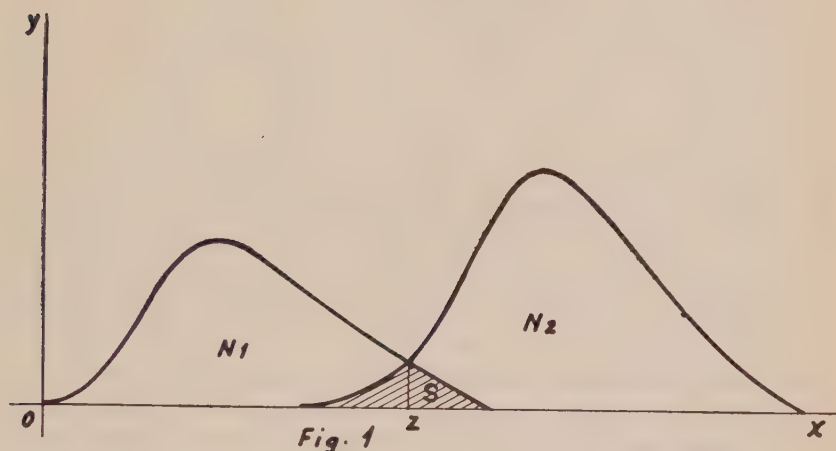
La integral de la (4.15), extendida a todo el intervalo de variabilidad de η , nos da la probabilidad de que en una realización del e.a. Λ , la v.a. η asuma un valor x perteneciente a la v.a. ξ_1 , si es $N_1 f_1(x)$ no mayor de $N_2 f_2(x)$ o perteneciente a la ξ_2 , si es $N_1 f_1(x)$ mayor de $N_2 f_2(x)$.

Simbolizando con C_1 a esta probabilidad, tenemos

$$C_1 = \frac{1}{N} \int_{-\infty}^{\infty} g(x) dx \quad (4.16)$$

donde la probabilidad contraria es $1 - C_1$, la cual tiene también un significado preciso, dado por la probabilidad de que en una realización del e.a. Λ , la v.a. η asuma un valor perteneciente a ξ_1 o a ξ_2 según sea $N_1 f_1(x)$ mayor o no mayor de $N_2 f_2(x)$.

Geométricamente, como se observa en la fig. 1, la probabilidad C_1 , es igual al área común comprendida entre las dos curvas de frecuencias y la abscisa, dividido por la suma de las áreas de las dos distribuciones



Simbolizando con S dicha área común, es, para el caso particular ilustrado en la fig. 1

$$C_1 = \frac{S}{N_1 + N_2} = \frac{S}{N} = \frac{N_2}{N} \int_{-\infty}^z f_2(x) dx + \frac{N_1}{N} \int_z^{\infty} f_1(x) dx \quad (4.16.a)$$

$$1 - C_1 = 1 - \frac{S}{N} = \frac{N_1}{N} \int_{-\infty}^z f_1(x) dx + \frac{N_2}{N} \int_z^{\infty} f_2(x) dx \quad (4.16.b)$$

El valor máximo que puede asumir C_1 , es no mayor de un medio. Es igual a un medio, en el caso particular en que sea $N_1 = N_2$ con ξ_1 y ξ_2 v.a. idénticas.

Definimos el área relativa de transvariación mediante el cociente entre el área de transvariación y su valor máximo, en la hipótesis $N_1 = N_2$ con ξ_1 y ξ_2 v.a. idénticas, es decir que es igual al área de transvariación dividida por un medio.

Simbolizándola con H_1 , tenemos

$$H_1 = 2 C_1 = \frac{2 S}{N_1 + N_2} \quad (4.17)$$

Para el caso particular $N_1 = N_2$; es

$$H_1 = 2 C_1 = \frac{S}{N_1} = \frac{S}{N_2} \quad (4.17.a)$$

que Gini (Ref.15) define relación de transvariación.

Entre el área relativa de transvariación y la probabilidad de transvariación, existe una gran afinidad, sin que en ningún momento pueda pensarse en una superposición de conceptos, desde que obedecen a definiciones independientes entre sí y completamente rigurosas. Señalamos sí, que a una probabilidad de transvariación igual a cero, corresponde también un área relativa de transvariación nula, no siendo exacta la afirmación recíproca, siempre que las *v.a.* consideradas sean discretas, como fácilmente se comprende y se puede ilustrar con ejemplos, como lo hacen Gini y Livada (Ref. 15).

En la tabla 1, damos un ejemplo de dos distribuciones con área relativa de transvariación igual a cero y probabilidad de transvariación distinta de cero.

TABLA 1

Variante	Frecuencias	
x	$N_1 f_1(x)$	$N_2 f_2(x)$
1	0	14
2	1	0
3	0	11
4	5	0
5	0	5
6	18	0

Si el conjunto de observaciones correspondiente a las *v.a.* ξ_1 y ξ_2 presentan el mismo intervalo de clase, el cálculo del área relativa de transvariación se obtiene en forma inmediata. Basta

sumar las frecuencias menores de cada intervalo, multiplicar por dos la suma y dividir el resultado por la suma de las frecuencias de las dos distribuciones.

Entre las características distintivas del área relativa de transvariación, señalamos

1) La función $g(x)$ es función continua de x , solamente cuando las funciones $f_1(x)$ y $f_2(x)$ que entran en su definición son continuas;

2) Si las funciones $f_1(x)$ y $f_2(x)$ son derivables, la función $g(x)$ también es derivable en todo su intervalo de continuidad, con excepción de los puntos de abscisa x que corresponda a los puntos de intersección de las funciones $N_1 f_1(x)$ y $N_2 f_2(x)$, pues en ellos la función $g(x)$ presenta puntos angulosos, no coincidiendo en consecuencia sus derivadas laterales en dichos puntos.

3) El área relativa de transvariación es igual a cero cuando las distribuciones $N_1 f_1(x)$ y $N_2 f_2(x)$ no presentan puntos de intersección, siendo en tal caso $g(x) = 0$ para todo valor de x ;

4) El área relativa de transvariación es igual a uno, cuando es $N_1 = N_2$ y las v.a. ξ_1 y ξ_2 son v.a. idénticas, verificándose en tal hipótesis que $N_1 f_1(x) = N_2 f_2(x) = g(x)$;

5) su medida es un número abstracto, de dimensión cero, por ello es apto para el estudio comparativo entre fenómenos cuyas observaciones están expresadas con diferentes unidades;

6) asume valores en el intervalo $[0,1]$;

7) El valor uno lo asume solamente en la hipótesis señalada en 4).

4.5. VALOR DISCRIMINATIVO

Tengamos, como en el párrafo anterior, dos colectivos de tamaño N_1 y N_2 respectivamente, sobre los cuales se realiza el e.a. Λ . Sean ξ_1 y ξ_2 sus correspondientes v.a. independientes, con medias $\alpha_{1.0} = M_1$ y $\alpha_{0.1} = M_2$ y f. de d. $f_1(x)$ y $f_2(x)$. Supongamos que

$$M_1 > M_2$$

El valor discriminativo es, por definición, el valor z de las v.a. ξ_1 y ξ_2 para el cual se verifica que es mínimo el error que se

comete admitiendo que todos los valores de la v.a. ξ_1 son mayores de z y todos los valores de la v.a. ξ_2 son menores de z .

Para un valor cualquiera x de ξ_1 y ξ_2 , tenemos, llamando $n_1(x)$ al error que se comete admitiendo que todos los valores de ξ_1 son mayores de x y $n_2(x)$ al error que se comete admitiendo que todos los valores de ξ_2 son menores de x

$$n_1(x) = N_1 \int_{-\infty}^x f_1(t) dt \quad (4.18)$$

$$n_2(x) = N_2 \int_x^{\infty} f_2(t) dt \quad (4.19)$$

Simbolizando con $\varepsilon(x)$ a la función del error que queremos hacer mínima, es

$$\varepsilon(x) = \frac{n_1(x) + n_2(x)}{N_1 + N_2} \quad (4.20)$$

de donde obtenemos

$$\frac{d\varepsilon(x)}{dx} = \frac{N_1 f_1(x) - N_2 f_2(x)}{N_1 + N_2}$$

siendo el valor discriminativo z igual a una de las raíces de la ecuación

$$N_1 f_1(x) - N_2 f_2(x) = 0 \quad (4.21)$$

que corresponde a la abscisa de uno de los puntos de intersección de las dos distribuciones. Si las raíces de la (4.21) corresponden a más de un mínimo relativo de la (4.20) la raíz que corresponda al mínimo absoluto, nos definirá el valor discriminativo z .

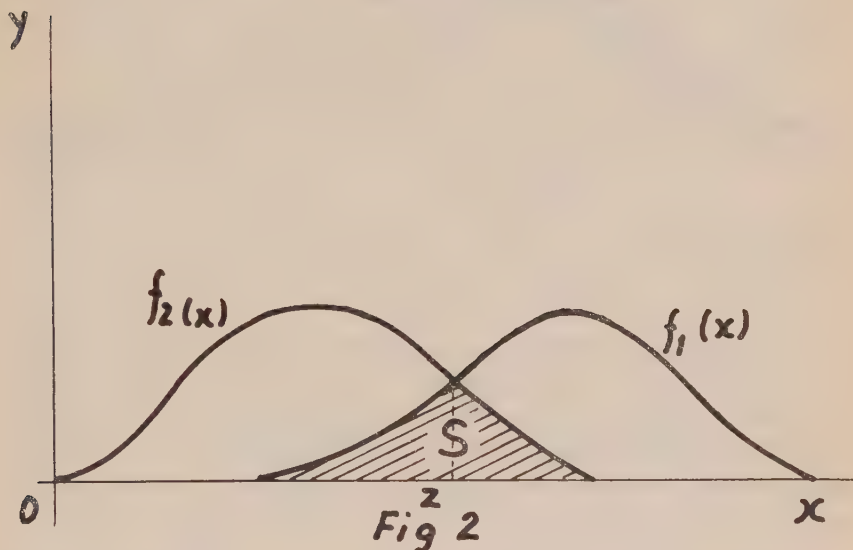
El error del valor discriminativo es

$$\varepsilon = \frac{n_1(z) + n_2(z)}{N_1 + N_2} \quad (4.22)$$

Para el caso particular que las dos distribuciones tengan el mismo número de observaciones, el error ε nos da una medida

de la tipicidad entre $f_1(x)$ y $f_2(x)$. Para tal supuesto, Gini y Livada (Ref. 15) con De Helguero, le dan al valor discriminativo el nombre de valor crítico.

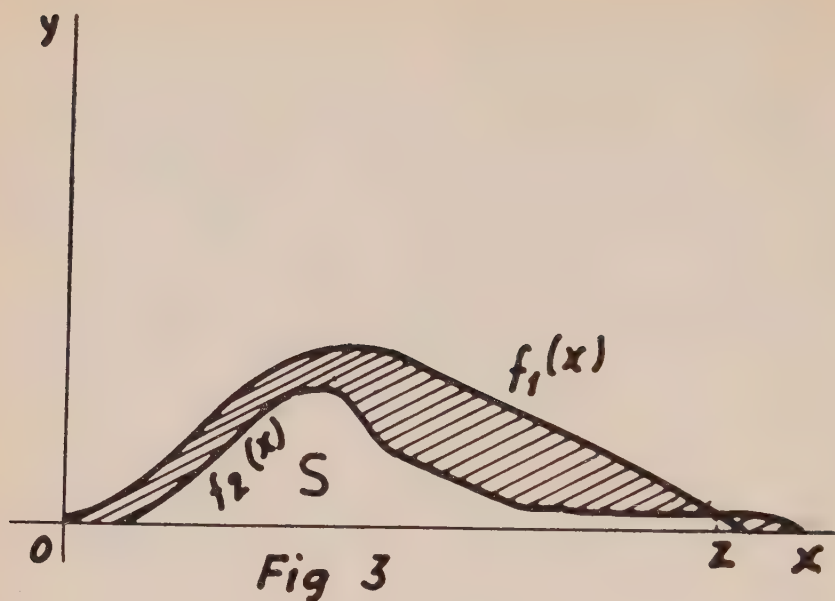
El error ϵ guarda con la probabilidad de transvariación y con el área relativa de transvariación, la independencia que impone la autonomía de su definición. Ello no impide que entre estos conceptos se encuentren analogías y afinidades. En efecto, si la (4.21) presenta una sola raíz y esta raíz se encuentra a la izquierda de M_1 , el error del valor discriminativo es igual a la mitad del área relativa de transvariación, como se ilustra gráficamente en la fig. 2; si la raíz se encuentra a la derecha de M_1 ,



el error ϵ es mayor que la mitad del área relativa de transvariación, siendo también mayor cuando las dos distribuciones presentan más de un punto de intersección, como se puede observar, respectivamente, en las figuras 3, 4 y 5. Es decir, que el error del valor discriminativo es no menor que la mitad del área relativa de transvariación.

Solamente para el caso ilustrado en la Fig 1, se presenta la igualdad entre el error ϵ y la mitad del área relativa de transvariación, por ser

$$S = n_1(z) + n_2(z)$$



mientras que para los restantes casos tenemos

$$S < n_1(z) + n_2(z)$$

y la diferencia está dada por la parte rayada de las Figs. 3, 4 y 5.

De los casos precedentemente considerados, el que más nos debe llamar la atención es el ilustrado en la Fig. 3. La magnitud de la diferencia entre el error del valor discriminativo y la mitad del área relativa de transvariación, es debida a la restricción impuesta a la definición del primero, restricción dada por la exigencia

$$M_1 > M_2$$

Si prescindimos de esta condición y buscamos el valor z para el cual la (4.20) presenta un mínimo, tomando alternativamente (véase Gini y Lívada, Ref. 15).

$$(I) \left\{ \begin{array}{l} n_1(x) = N_1 \int_{-\infty}^x f_1(t) dt \\ n_2(x) = N_2 \int_x^{\infty} f_2(t) dt \end{array} \right. \quad (4.23)$$

$$(II) \left\{ \begin{array}{l} n_1(x) = N_1 \int_x^{\infty} f_1(t) dt \\ n_2(x) = N_2 \int_{-\infty}^x f_2(t) dt \end{array} \right. \quad (4.24)$$

tendremos entonces dos mínimos, correspondientes cada uno, respectivamente, a las hipótesis (I) y (II). De estas dos soluciones, el valor de z que dé un menor valor de ϵ nos definirá el valor discriminativo. Con este criterio encontramos que también para el caso ilustrado en la Fig. 3, es

$$S = n_1(z) + n_2(z)$$

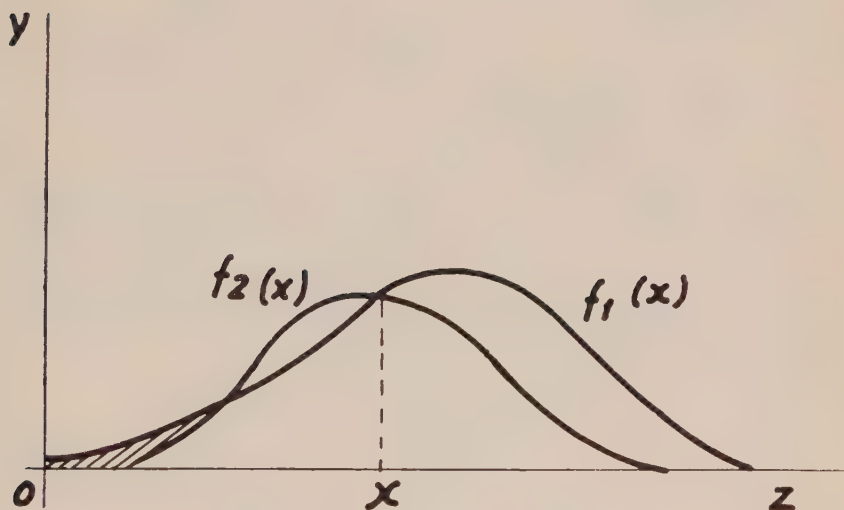


Fig. 4

A partir de este último punto de vista dado para la definición del valor discriminativo, concluimos que, si las dos distribuciones presentan un solo punto de intersección, el error del valor discriminativo es igual a la mitad del área relativa de transvaciación; en los restantes casos sigue siendo mayor.

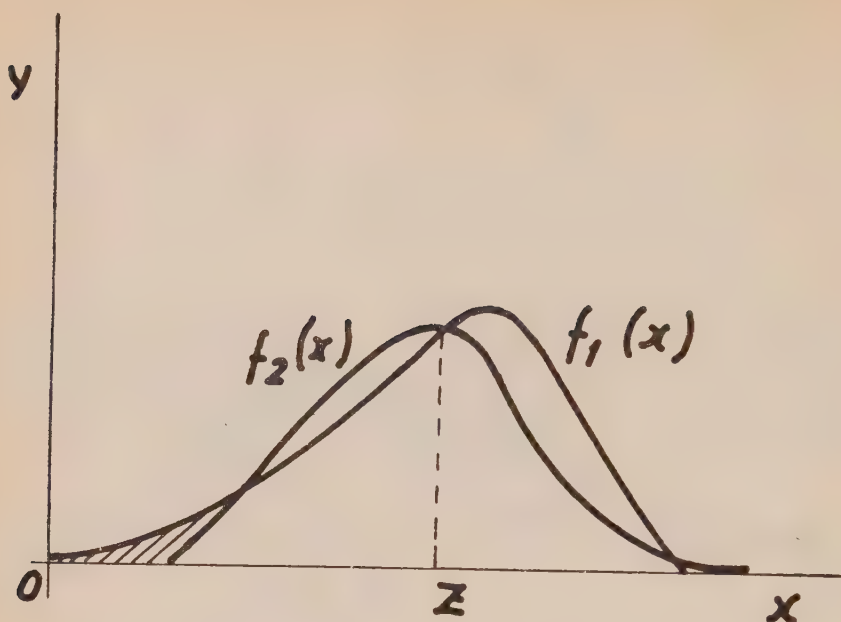


Fig. 5

4.6. TRANSVARIACIÓN EN LA HIPOTESIS DE DISTRIBUCIÓN NORMAL.

Deducimos a continuación los parámetros de la transvariación, en la hipótesis que las v.a. ξ_1 y ξ_2 siendo independientes, tengan una distribución normal, con medias

$$\begin{cases} \alpha_{1.0} = M_1 \\ \alpha_{0.1} = M_2 \end{cases}$$

y varianzas

$$\begin{cases} \sigma_1^2 = \mu_{2.0} \\ \sigma_2^2 = \mu_{0.2} \end{cases}$$

Partimos del supuesto que $M_1 > M_2$

4.6.I. - Probabilidad e intensidad de transvariación

Evidentemente estamos en presencia de un caso particular del considerado en 3.9, dado por la condición de independencia,

lo que equivale a decir que las v.a. ξ_1 y ξ_2 se encuentran incorrelacionadas, o sea que

$$\rho = 0$$

En consecuencia, a partir respectivamente de (3.46) y (3.47), o lo que es lo mismo, a partir de (3.56) y (3.57), luego de hacer $\rho = 0$, tenemos la probabilidad de transvariación y la intensidad de transvariación de orden r , en la hipótesis de dos variables aleatorias independientes y normalmente distribuidas, siendo

$$\left\{ \begin{array}{l} \alpha = M_2 - M_1 = -d < 0 \\ \sigma^2 = \sigma_1^2 + \sigma_2^2 \\ d_1 = \frac{d}{\sqrt{\sigma_1^2 + \sigma_2^2}} = -\frac{\alpha}{\sigma} > 0 \end{array} \right. \quad (4.25)$$

4.6.2. Área relativa de transvariación

Siendo respectivamente N_1 y N_2 el número de elementos de cada uno de los colectivos considerados, pasamos a determinar los puntos de intersección de las dos distribuciones normales, lo que nos permitirá calcular el área común y con ella, el área relativa de transvariación.

Los puntos de intersección están dados por los valores x de las v.a. ξ_1 y ξ_2 que satisfagan la igualdad

$$\frac{N_1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-M_1)^2}{2\sigma_1^2}} = \frac{N_2}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-M_2)^2}{2\sigma_2^2}} \quad (4.26)$$

de donde, logaritmando

$$\begin{aligned} (\sigma_2^2 - \sigma_1^2) x^2 - 2(\sigma_2^2 M_1 - \sigma_1^2 M_2) x + (\sigma_2^2 M_1^2 - \sigma_1^2 M_2^2) + \\ + 2\sigma_1^2 \sigma_2^2 \log \frac{\sigma_1}{\sigma_2} \cdot \frac{N_2}{N_1} = 0 \end{aligned} \quad (4.27)$$

$$x = \frac{(\sigma_2^2 M_1 - \sigma_1^2 M_2) \pm \sigma_1 \sigma_2 \sqrt{(M_1 - M_2)^2 - 4,6052 (\sigma_2^2 - \sigma_1^2) \log \frac{\sigma_1}{\sigma_2} \cdot \frac{N_2}{N_1}}}{\sigma_2^2 - \sigma_1^2} \quad (4.28)$$

estando x expresado en (4.28) en función de los logaritmos decimales.

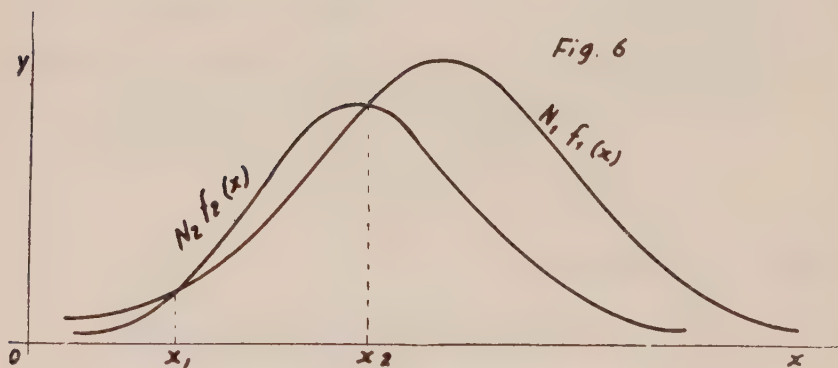
Podemos analizar los siguientes casos, con respecto a las raíces que verifican la (4.26). Con Δ , simbolizamos el discriminante de la (4.27).

a) $\sigma_1 \neq \sigma_2$; $\Delta > 0$: Para este caso tendremos siempre dos raíces reales y distintas. Si se verifica simultáneamente la condición

$$\begin{cases} \sigma_1 \leq \sigma_2 \\ \sigma_1 N_2 \leq \sigma_2 N_1 \end{cases} \quad (4.29)$$

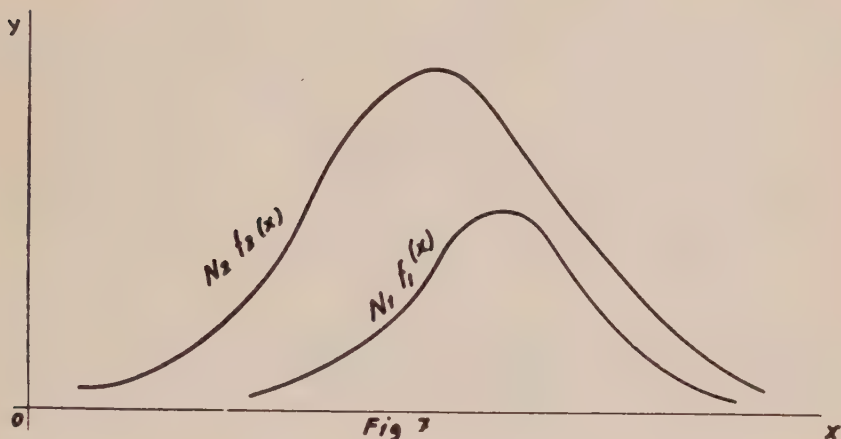
el discriminante será igual a la suma de dos términos positivos, dado que el segundo tiene los factores $(\sigma_2^2 - \sigma_1^2)$ y $\log \frac{\sigma_1}{\sigma_2} \cdot \frac{N_2}{N_1}$ con signos distintos. Si se verifica simultáneamente la condición

$$\begin{cases} \sigma_1 \leq \sigma_2 \\ \sigma_1 N_2 \geq \sigma_2 N_1 \end{cases} \quad (4.30)$$



El segundo término del discriminante es negativo, luego será suficiente, para que $\Delta > 0$, que dicho término sea, en valor absoluto, menor que el primero. Gráficamente lo ilustramos con la Fig. 6.

b) $\sigma_1 \neq \sigma_2$; $\Delta < 0$: Para este caso tendremos siempre dos raíces complejas conjugadas. Para ello es condición necesaria pero no suficiente, que se verifique la (4.30). Decimos que no es suficiente, porque dado que con la condición (4.30) el segundo término del discriminante es negativo, deberá verificarse además que, en valor absoluto, sea mayor que el primero.



La existencia de raíces complejas conjugadas no significa otra cosa que las dos distribuciones no presentan puntos de intersección, en consecuencia, como ambas distribuciones están definidas para todo valor real de sus respectivas variables aleatorias, una de ellas tendrá frecuencias permanentemente menores a la otra, es decir que, gráficamente, estará íntegramente contenida por ésta. Suponiendo $N_1 < N_2$, tendremos para cualquier valor real de x

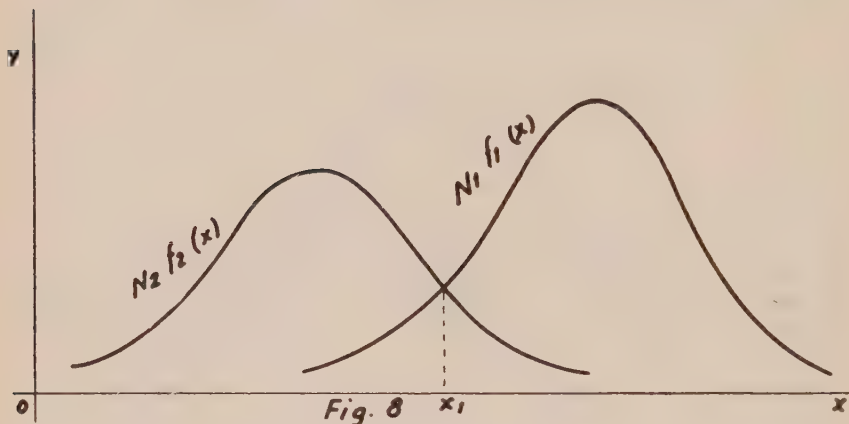
$$\frac{N_1}{\sigma_1 \sqrt{2\pi}} e^{-\frac{(x-M_1)^2}{2\sigma_1^2}} < \frac{N_2}{\sigma_2 \sqrt{2\pi}} e^{-\frac{(x-M_2)^2}{2\sigma_2^2}}$$

Gráficamente lo ilustramos con la Fig. 7.

c) $\sigma_1 \neq \sigma_2$; $N_1 = N_2$: El presente constituye un caso particular del considerando en a), en consecuencia tenemos, como en a), dos raíces reales y distintas.

d) $\sigma_1 = \sigma_2$; $N_1 \neq N_2$; $M_1 \neq M_2$: En este caso, las dos distribuciones normales presentan un solo punto de intersección, que está dada por la raíz de la ecuación lineal

$$2 (M_1 - M_2) x - (M_1^2 - M_2^2) - 2 \sigma_1^2 \log \frac{N_2}{N_1} = 0 \quad (4.31)$$



obtenida a partir de la (4.27). Gráficamente lo ilustramos con la Fig. 8.

e) $\sigma_1 = \sigma_2$; $N_1 = N_2$; $M_1 \neq M_2$: El presente constituye un caso particular del anterior, con la característica de que la abscisa del punto de intersección es igual a la semisuma de las medias aritméticas de las distribuciones consideradas. A partir de (4.31) luego de hacer $N_1 = N_2$, obtenemos

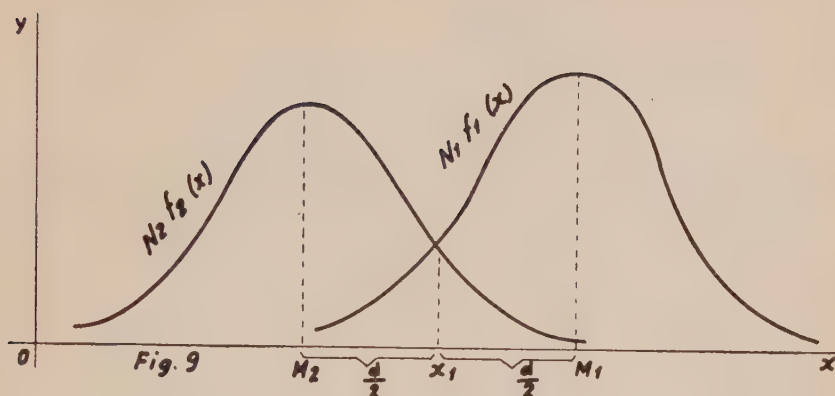
$$x_1 = \frac{M_1 + M_2}{2}$$

Gráficamente lo ilustramos con la Fig. 9

f) $\sigma_1 = \sigma_2$; $N_1 \neq N_2$; $M_1 = M_2$: Las dos distribuciones no presentan punto de intersección, teniendo una de ellas frecuencias permanentemente menores a la otra, con la característica de que su cociente es constantemente igual al cociente de N_1 y N_2 .

g) $\sigma_1 = \sigma_2$; $N_1 = N_2$; $M_1 = M_2$: El presente es un caso trivial; corresponde al supuesto de variables idénticas.

$$x_1 = \frac{M_1 + M_2}{2}$$



A continuación obtenemos el área relativa de transvariación para cada uno de los casos considerados precedentemente.

Caso a) :

$$H_1 = \frac{2S}{N_1 + N_2} = \frac{2}{N_1 + N_2} [N_2(1 + F_2(x_1) - F_2(x_2)) + N_1(F_1(x_2) - F_1(x_1))] \quad (4.32)$$

Caso b) :

$$H_1 = \frac{2S}{N_1 + N_2} = \frac{2N_1}{N_1 + N_2} \quad (4.33)$$

Caso c) :

$$H_1 = \frac{2S}{N_1 + N_2} = 1 + F_2(x_1) + F_1(x_2) - F_2(x_2) - F_1(x_1) \quad (4.34)$$

Caso d) :

$$H_1 = \frac{2S}{N_1 + N_2} = \frac{2}{N_1 + N_2} [N_1 F_1(x_1) + N_2(1 - F_2(x_1))] \quad (4.35)$$

Caso e) :

$$H_1 = \frac{2S}{N_1 + N_2} = 1 + F_1(x_1) - F_2(x_1) = 2\Phi_1(z_1) \quad (4.36)$$

resultado al que llegamos mediante sencillas consideraciones lógicas, o a través de las siguientes sustituciones

$$z = \frac{x - M_1}{\sigma_1} \quad \text{en} \quad F_1(x_1)$$

$$z = \frac{x - M_2}{\sigma_1} \quad \text{en} \quad F_2(x_1)$$

Recordando que

$$x_1 = \frac{M_1 + M_2}{2}$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2$$

obtenemos

$$z_1 = \frac{M_2 - M_1}{2 \sigma_1} = -\frac{d}{2 \sigma_1} = -\frac{d}{\sigma \sqrt{2}}$$

$$z_2 = \frac{M_1 - M_2}{2 \sigma_1} = \frac{d}{\sigma \sqrt{2}} = -z_1$$

Caso f) :

Si es $N_1 < N_2$, tenemos

$$H_1 = \frac{2 S}{N_1 + N_2} = \frac{2 N_1}{N_1 + N_2} \quad (4.37)$$

Caso g) :

$$H_1 = \frac{2 S}{N_1 + N_2} = 1$$

Observamos que para ninguna de las hipótesis consideradas, hemos obtenido un área relativa de transvariación nula. Ello es evidente, porque la condición necesaria y suficiente para que dicha área sea nula, es que las dos distribuciones no presenten área común, situación de presentación imposible para el caso de distribuciones normales, por el hecho de que las mismas están definidas para todo valor real de sus respectivas variables aleatorias.

De acuerdo con la notación que venimos empleando cuando trabajamos con distribuciones normales, recordamos que

$$F(x) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-M)^2}{2\sigma^2}} dt$$

es decir que con la F , simbolizamos la f. de p. a. de la v.a. normal (M, σ) , mientras que con Φ , simbolizamos la f. de p.a. de la v.a. normal $(0,1)$, es decir que

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

4.6.3. Valor discriminativo

En la determinación del valor discriminativo para el caso particular de dos distribuciones normales, se pueden presentar solo tres alternativas posibles, a saber :

- a) las dos distribuciones no presentan punto de intersección ;
- b) las dos distribuciones presentan un solo punto de intersección ;
- c) las dos distribuciones presentan dos puntos de intersección.

a) Para la primera hipótesis, la determinación del valor discriminativo carece en absoluto de interés, porque ello nos llevaría a los valores extremos de la variable aleatoria, es decir a $+\infty$ y a $-\infty$, donde ambas distribuciones son asintóticamente nulas. Aceptando el análisis extremo y suponiendo $N_1 < N_2$, tenemos para el error del valor discriminativo, en base a las consideraciones que nos condujeron a (4.23) y (4.24).

$$\varepsilon() = \varepsilon(-\infty) = \infty + \frac{N_1}{N_1 + N_2}$$

que es igual a la mitad del área relativa de transvariación, como se desprende comparando este resultado con (4.33) y (4.37).

b) Cuando las dos distribuciones presentan un solo punto de intersección, vimos que el valor discriminativo es igual a la abscisa de dicho punto. Para el caso de dos distribuciones normales, el valor discriminativo es igual a la raíz de la (4.31), es decir

$$x_1 = \frac{M_1 + M_2}{2} + \frac{\sigma_1^2}{M_1 - M_2} \log \frac{N_2}{N_1} \quad (4.38)$$

El error del valor discriminativo es

$$\epsilon(x_1) = \frac{n_1(x_1) + n_2(x_1)}{N_1 + N_2} = \frac{N_1 F_1(x_1) + N_2 [1 - F_2(x_1)]}{N_1 + N_2} \quad (4.39)$$

que es igual a la mitad del área relativa de transvariación, cuyo valor está dado por (4.35).

Para el caso particular de ser $N_1 = N_2$ tenemos para la

$$x_1 = \frac{M_1 + M_2}{2} \quad (4.38)$$

que recibe el nombre de valor crítico. El valor del error ϵ , es

$$\epsilon(x_1) = \frac{1 + F_1(x_1) - F_2(x_1)}{2} = \Phi_1(z_1) \quad (4.40)$$

que es igual a la mitad de la realación de transvariación, como se desprende de la (4.36), donde se dan las sustituciones realizadas.

A partir de (3.56) y (4.25), obtenemos para la probabilidad de transvariación

$$P = 2 \Phi_1(-d_1) = 2 \Phi_1\left(-\frac{d}{\sigma}\right) = 2 \Phi_1(z_1 \sqrt{2}) \quad (4.41)$$

con lo que, siendo Z_1 , una cantidad negativa, por la condición $M_1 > M_2$, deducimos que la relación de transvariación es mayor que la probabilidad de transvariación.

c) Si las dos distribuciones presentan dos puntos de intersección, el valor discriminativo es igual al punto que haga mínimo el error ϵ . Simbolizando con x_2 al valor discriminativo, tenemos

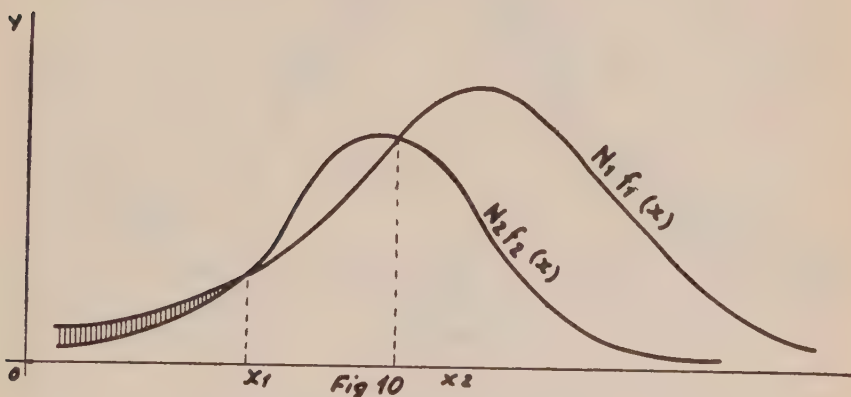
para el error ϵ , a partir de la (4.23), que es la que nos da, en este caso particular, véase Fig. 10, el valor mínimo de ϵ :

$$\epsilon = \frac{n_1(x_2) + n_2(x_2)}{N_1 + N_2} = \frac{N_1 F_1(x_2) + N_2 [1 - F_2(x_2)]}{N_1 + N_2} \quad (4.42)$$

valor que es mayor que la mitad del área relativa de transvariación en una cantidad igual a

$$\epsilon - \frac{H_1}{2} = \frac{N_1 F_1(x_1) - N_2 F_2(x_1)}{N_1 + N_2} \quad (4.42.a)$$

como se deduce en forma inmediata, a partir de (4.42) y (4.32). Su numerador viene dado, gráficamente, por la parte rayada de la Fig. 10. Si tomamos como valor discriminativo la raíz x , o



trabajamos, para cualquiera de las dos raíces, en base a la hipótesis (4.24), el valor de ϵ será mayor, como fácilmente se puede deducir de la Fig. 10.

Para el caso particular ilustrado en c) de 4.6.2., tenemos

$$\epsilon = \frac{1 + F_1(x_2) - F_2(x_2)}{2} \quad (4.43)$$

A partir de (4.43) y (4.34), deducimos que el error del valor discriminativo excede a la mitad del área relativa de transvariación en la cantidad

$$\varepsilon - \frac{H_1}{2} = \frac{F_1(x_1) - F_2(x_1)}{2} \quad (4.43.a)$$

resultado que se obtiene igualmente a partir de (4.42.a), luego de hacer $N_1 = N_2$.

Podemos aún considerar un interesante caso particular dentro de este caso particular, dado por la hipótesis adicional de que

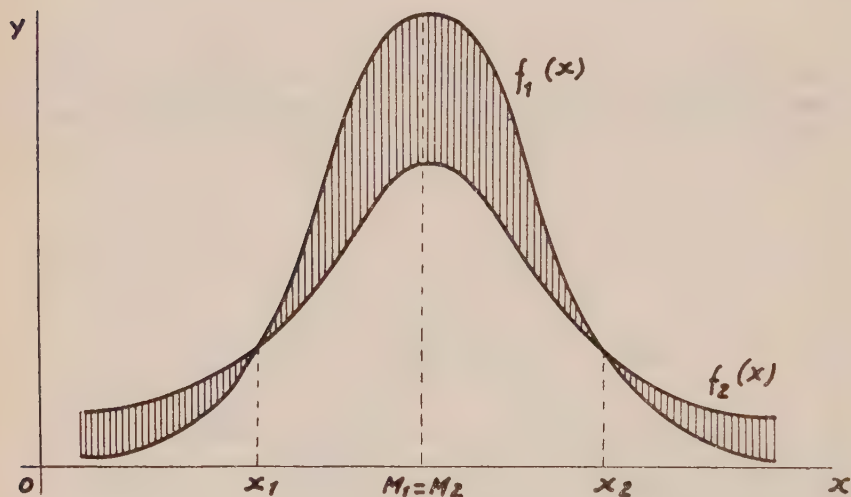


Fig. 11

las medias de las dos distribuciones normales sean iguales, es decir que $M_1 = M_2$.

La (4.27) se transforma en

$$(\sigma_2^2 - \sigma_1^2) x^2 - 2 M_1 (\sigma_2^2 - \sigma_1^2) x + M_1^2 (\sigma_2^2 - \sigma_1^2) +$$

$$+ 2 \sigma_1^2 \sigma_2^2 \log \frac{\sigma_1}{\sigma_2} = 0$$

siendo

$$\begin{aligned}
 x &= M_1 \mp \sigma_1 \sigma_2 \sqrt{2 \frac{\log \sigma_2 - \log \sigma_1}{\sigma_2^2 - \sigma_1^2}} = \\
 &= M_1 \mp \sigma_1 \sigma_2 \sqrt{4,6052 \frac{\log_{10} \sigma_2 - \log_{10} \sigma_1}{\sigma_2^2 - \sigma_1^2}}
 \end{aligned} \quad (4.44)$$

donde el segundo miembro lo expresamos en función de los logaritmos naturales y el tercero en función de los logaritmos decimales.

Las abscisas de los dos puntos de intersección, dadas por (4.44), son simétricas del punto M_1 , dándonos el mismo valor para el error ε , con lo que cualquiera de las dos raíces puede ser considerada como valor crítico. Esta situación vuelve no significativo dichos valores, lo que cabía esperar dada la igualdad de las medias. En efecto, tomando como valor crítico x_2 , para el cual debemos trabajar con la hipótesis (4.24), tenemos

$$\varepsilon_1(x_2) = \frac{1 - F_1(x_2) + F_2(x_2)}{2} = \frac{F_1(x_1) + F_2(x_2)}{2} < \frac{1}{2} \quad (4.45)$$

y tomando como valor crítico x_1 , y trabajando ahora con la hipótesis (4.23), tenemos

$$\varepsilon_1(x_1) = \frac{F_1(x_1) + 1 - F_2(x_1)}{2} = \frac{F_1(x_1) + F_2(x_2)}{2} < \frac{1}{2} \quad (4.45.a)$$

dándonos el mismo valor para el error ε . Las otras combinaciones entre las raíces x_1 y x_2 y las hipótesis (4.23) y (4.24), nos dan valores mayores para ε . Tales valores son

$$\varepsilon_2(x_1) = \varepsilon_2(x_2) = \frac{F_1(x_2) + F_2(x_1)}{2} > \frac{1}{2}$$

como fácilmente se puede ver en la Fig. 11. Para ello debemos recordar que estamos trabajando con $N_1 = N_2$ y con $M_1 = M_2$, razón por la cual, del área unidad que encierra cada distribución con la abscisa, Fig. 11, el área de $f_1(x)$ que excede a $f_2(x)$, es igual a la suma de las dos áreas de las ramas extremas donde $f_2(x)$ excede a $F_1(x)$, siendo a su vez iguales estas dos áreas extremas, en virtud de la simetría de la distribución con respecto

a su media y de la simetría de las raíces con respecto a M_1 . Estas áreas se encuentran rayadas en el gráfico.

La probabilidad de transvariación es igual a la unidad. En efecto, a partir de (3.56) y (4.25), unida a la condición $M_1 = M_2$, tenemos

$$P = 2 \Phi(0) = 1$$

La relación de transvariación es, recordando su definición

$$H_1 = 1 + F_1(x_1) + F_2(x_2) - F_1(x_2) - F_2(x_1)$$

de donde, teniendo presente la ubicación relativa de x_1 y x_2

$$H_1 = 1 + 2 F_1(x_1) - 2 F_2(x_1)$$

que expresada en función de la variable aleatoria standardizada, nos da, teniendo presente la (4.44)

$$H_1(u) = 1 + 2 \Phi\left(-\sqrt{2 \frac{\log u}{u^2 - 1}}\right) - 2 \Phi\left(-\sqrt{2 u^2 \frac{\log u}{u^2 - 1}}\right) \quad (4.46)$$

siendo

$$u = \frac{\sigma_1}{\sigma_2} < 1 \quad (4.47)$$

La (4.46) toma valores de 0 a 1, para u variando también de 0 a 1. Luego de resolver la indeterminación que se nos presenta para estos valores extremos de u , tenemos

$$H_1(0) = 1 + 2 \Phi(-\infty) - 2 \Phi(0) = 0$$

$$H_1(1) = 1 + 2 \Phi(-1) - 2 \Phi(-1) = 1$$

El caso límite que se presenta para cuando u tiende a cero, corresponde no ya a la transvariación entre dos distribuciones normales, sino entre una distribución normal y la distribución $\omega(x)$, cuya f. de d. es igual a 1 para $x = M_1$ y es igual a cero para todo otro valor de x . El área común es evidentemente nula. El estudio de la transvariación en R_1 , puede fácilmente asimilarse al estudio de la transvariación entre dicha distribución en R_1 , y la función de distribución $\omega(x)$.

5. — TRANSVARIACION EN EL ESPACIO R_n

5.1. INTRODUCCIÓN

En los capítulos precedentes nos hemos ocupado de la transvariación en el espacio unidimensional y bidimensional. En el primer caso estudiamos la transvariación entre la v.a. unidimensional ξ y los parámetros arbitrarios A y λ ; en el segundo nos ocupamos de la transvariación entre la v.a. bidimensional $\xi = (\xi_1, \xi_2)$ y el parámetro $\lambda = (\lambda_1, \lambda_2)$. Para esta v.a. bidimensional estudiamos sucesivamente la transvariación en la hipótesis que sus variables componentes se encuentren correlacionadas, normalmente correlacionadas, sean independientes y su caso particular de variables aleatorias normalmente independientes. En 3.1., hemos definido la transvariación entre la v.a. bidimensional $\xi = (\xi_1, \xi_2)$ y el parámetro $\lambda = (\lambda_1, \lambda_2)$ que es, en síntesis, la transvariación entre dos v.a. unidimensionales, ξ_1 y ξ_2 , con respecto a sus correspondientes parámetros λ_1 y λ_2 .

Siguiendo un ordenamiento lógico llegaremos al caso general de la transvariación entre dos v.a. n -dimensionales, con respecto a un determinado parámetro de cada una de sus $2n$ componentes unidimensionales, para $n = 1, 2, 3, \dots$, con un número finito o a lo sumo, teóricamente, una cantidad numerable de dimensiones. A continuación nos ocupamos de la transvariación entre una v.a. n -dimensional $\xi = (\xi_1, \xi_2, \dots, \xi_n)$ y el parámetro n -dimensional $A = (A_1, A_2, \dots, A_n)$ con respecto al parámetro también n -dimensional $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$.

El caso particular $n = 1$, fué objeto de estudio en el capítulo 2. Dado el e.a. Λ , al cual asociamos la v.a. ξ , que puede asumir valores en el espacio R_n , siendo $x = (x_1, x_2, \dots, x_n)$, una particular determinación de la v.a. $\xi = (\xi_1, \xi_2, \dots, \xi_n)$, cuya f. de d. es $f(x_1, x_2, \dots, x_n)$, tenemos para su f. de p.a.

$$P(\xi \leq x) = F(x) = P(\xi_1 \leq x_1, \dots, \xi_n \leq x_n) = \int_S dP(s) \quad (5.1)$$

siendo S el subespacio de R_n , definido por la relación $\xi \leq x$.

Siendo

$$\bar{S} = R_n - S$$

tenemos

$$P(\bar{s}) = P(R_n - S) = P(R_n) - P(S) = 1 - P(s)$$

Si consideramos un grupo de $k < n$ componentes de la v.a. ξ , mientras hacemos tender a $+\infty$ las restantes $n - k$ componentes, tendremos definida la distribución marginal k -dimensional de la v.a. ξ , con respecto a dichas k componentes. Esta distribución marginal de probabilidad puede ser interpretada como el resultado de proyectar la masa unitaria de la distribución de la v.a. n -dimensional sobre el subespacio dado por las k componentes consideradas.

La f. de p.a. de la distribución marginal k -dimensional, para las k primeras componentes, es

$$\begin{aligned} F(x_1, x_2, \dots, x_k) &= F(x_1, x_2, \dots, x_k, +\infty, \dots, +\infty) = \\ &= P(\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_k \leq x_k) \end{aligned} \quad (5.1.a)$$

Si las componentes $\xi_1, \xi_2, \dots, \xi_n$, de la v.a. n -dimensional ξ , son v.a. independientes, tenemos

$$F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i) = \prod_{i=1}^n P(\xi_i \leq x_i) \quad (5.1.b)$$

siendo, para el caso de distribución marginal k -dimensional

$$F(x_1, x_2, \dots, x_k) = \prod_{i=1}^k F_i(x_i) = \prod_{i=1}^k P(\xi_i \leq x_i) \quad (5.1.c)$$

Introduciendo los parámetros $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$ y $A = (A_1, A_2, \dots, A_n)$, por ahora arbitrarios, nos definirá una transvariación marginal k -dimensional, para $k < n$, entre la v.a. ξ y el parámetro A , con respecto al parámetro λ , en una realización determinada del e.a. Λ , cuando el signo de la diferencia entre el valor x de la v.a. ξ y el parámetro A , sea opuesto, para k componentes de la v.a. ξ , al signo de la diferencia entre sus correspondientes k componentes de los parámetros λ y A . Considerando las

k primeras componentes, tendremos, por definición, una transvariación marginal k -dimensional, cuando en la realización del e.a. Λ , la *v.a.* ξ presenta un valor x tal que, el signo de la diferencia $x_i - A_i$, sea opuesto al signo de la diferencia $\lambda_i - A_i$, simultáneamente para $i = 1, 2, \dots, k$. El valor absoluto de la potencia

$$\prod_{i=1}^k (x_i - A_i)^r \quad (5.2)$$

nos definirá su respectiva intensidad de orden r . El valor absoluto de la (5.2), para r igual a uno, nos define, al mismo tiempo que la intensidad de orden uno, la medida de dicha transvariación.

La medida de transvariación de mayor valor absoluto, de entre todas las posibles transvariaciones marginales k -dimensionales, nos definirá el campo de transvariación marginal k -dimensional, entre la *v.a.* ξ y el parámetro A , con respecto al parámetro λ .

Si nos encontramos con s_1 casos para los cuales sea nula una de las k diferencias $x_i - A_i$, para $i = 1, 2, \dots, k$, siendo $\lambda_i \neq A_i$, convendremos en considerar la mitad de ellos como teniendo signo contrario al signo de la diferencia $\lambda_i - A_i$. En general, simbolizando con s_j , siendo $j \leq k$, el número de casos para los cuales sean nulas j de las k diferencias $x_i - A_i$, siendo $\lambda_i \neq A_i$

convendremos en considerar $\frac{s_j}{2^j}$ de ellos como teniendo signo contrario al signo de la diferencia $\lambda_i - A_i$. Esta convención obedece al siguiente razonamiento: suponemos que cada uno de los s_j casos tiene 2^j formas igualmente posibles de ser consideradas para la transvariación, de las cuales $\binom{j}{h}$, $0 \leq h \leq j$ formas corresponden a la hipótesis que sean nulas h de las j diferencias, de

donde $\frac{s_j \binom{j}{h}}{2^j}$ es el número de casos, del total s_j , que consideramos con h diferencias positivas; para $h = j$, tenemos $\frac{s_j}{2^j}$ casos, los cuales consideramos que constituyen transvariaciones marginales k -dimensionales.

Si en cambio es $\lambda_i = A_i$, para h de los k valores de i , convendremos en considerar la transvariación marginal k -dimensional

como una transvariación marginal de $k - h$ dimensiones, respecto a las $k - h$ v.a. para las cuales $\lambda_i \neq A_i$.

5.2. PROBABILIDAD DE TRANSVARIACIÓN MARGINAL K- DIMENSIONAL

Antes de definir este concepto, definiremos sus elementos componentes, que son la transvariabilidad marginal k -dimensional y su valor máximo.

Definimos la transvariabilidad marginal k -dimensional, entre la v.a.

$$\xi = (\xi_1, \xi_2, \dots, \xi_n)$$

y el parámetro

$$A = (A_1, A_2, \dots, A_n)$$

con respecto al parámetro

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n)$$

a la probabilidad de que en una realización determinada del e.a. Λ , la v.a. ξ asuma un valor x tal que la diferencia $x_i - A_i$, para k determinadas componentes de la v.a. ξ , sea de signo opuesto al signo de la diferencia $\lambda_i - A_i$, para esas mismas componentes.

Sin pérdida de generalidad supondremos en todo lo que sigue que es $\lambda_i > A_i$, para $i = 1, 2, \dots, n$.

Considerando las k primeras componentes de la v.a. ξ , tenemos, por definición de transvariabilidad marginal k -dimensional.

$$p_{n,k} = P(\xi_1 < A_1, \dots, \xi_k < A_k) = F(A_1, A_2, \dots, A_k) = \quad (5.3)$$

$$= \int_{-\infty}^{A_1} dx_1 \int_{-\infty}^{A_2} dx_2 \dots \int_{-\infty}^{A_k} f(x_1, x_2, \dots, x_k) dx_k$$

que es igual a un espacio de k dimensiones, parte del espacio unidad, limitado por las coordenadas A_1, A_2, \dots, A_k del hiperplano respectivo.

Introduciendo el parámetro real arbitrario

$$t = (t_1, t_2, \dots, t_n)$$

expresamos la (5.3) en función de las k primeras componentes de este parámetro, haciendo

$$\begin{aligned} p_{n,k}(t) &= P(\xi_1 < A_1 + t_1, \dots, \xi_k < A_k + t_k) = \\ &= F(A_1 + t_1, \dots, A_k + t_k) \end{aligned} \quad (5.4)$$

lo que nos define también una transvariabilidad marginal k -dimensional.

El máximo de (5.4), por lógica consecuencia de las propiedades de toda distribución de probabilidad de cualquier número de dimensiones, se obtiene para $t_i = \lambda_i - A_i$, luego

$$p_{n,k;M} = P(\xi_1 < \lambda_1, \dots, \xi_k < \lambda_k) = F(\lambda_1, \dots, \lambda_k) \quad (5.5)$$

Si λ_i es igual al valor mediano de la v.a. ξ_i para $i = 1, 2, \dots, k$, entonces es $p_{n,k;M} = \frac{1}{2}$, por definición de hiperplano mediano. El máximo de la (5.5) es igual a uno, valor que alcanza cuando $\lambda_i \rightarrow +\infty$ simultáneamente para $i = 1, 2, \dots, k$. Para este caso límite, tenemos

$$p_{n,k;M} = F(+\infty, \dots, +\infty) = 1$$

El mínimo de la transvariabilidad marginal k -dimensional es igual a cero, para lo que es suficiente que $A_i \rightarrow -\infty$, para una y sólo una de las k componentes, es decir

$$p_{n,k} = F(-\infty, A_2, \dots, A_k) = \dots = F(A_1, \dots, A_{k-1}, -\infty) = 0$$

Definimos la probabilidad de transvariación marginal k -dimensional, considerando sus k primeras componentes, mediante el cociente entre la transvariabilidad marginal k -dimensional, de estas k primeras componentes, y su valor máximo. Luego, por definición es

$$P_{n,k} = \frac{p_{n,k}}{p_{n,k;M}} = \frac{F(A_1, A_2, \dots, A_k)}{F(\lambda_1, \lambda_2, \dots, \lambda_k)} \quad (5.6)$$

Dando a k valores desde 1 a $(n-1)$, tenemos respectivamente la probabilidad de transvariación marginal de $1, 2, \dots, (n-1)$

dimensiones. Considerando la componente ξ_i de la v.a. ξ , tenemos para la probabilidad de transvariación marginal unidimensional.

$$P_{n,1} = \frac{p_{n,1}}{p_{n,1,M}} = \frac{F(+\infty, \dots, +\infty, A_i, +\infty, \dots, +\infty)}{F(+\infty, \dots, +\infty, \lambda_i, +\infty, \dots, +\infty)} = \frac{F(A_i)}{F(\lambda_i)} \quad (5.6.a)$$

que es igual a la obtenida en (2.7).

En consecuencia, lo tratado en aquel capítulo puede ser considerado dentro de este esquema más general, con el título de teoría de la transvariación marginal unidimensional. Las características de la probabilidad de transvariación unidimensional dadas en 2.4, se pueden extender sin ninguna complicación, al presente caso de probabilidad de transvariación marginal k -dimensional.

Para $k = n$, tendremos, por definición, la probabilidad de transvariación n -dimensional. Deja ya, evidentemente, de ser probabilidad marginal, desde que consideramos todas sus componentes. Para su expresión analítica, tenemos

$$P_n = \frac{p_n}{p_{n,M}} = \frac{F(A_1, A_2, \dots, A_n)}{F(\lambda_1, \lambda_2, \dots, \lambda_n)} \quad (5.6.b)$$

Considerando el espacio hilbertiano, compuesto de un conjunto infinito numerable de dimensiones, desaparece este caso límite, pues siempre estaríamos en presencia de probabilidades marginales de transvariación.

5.3. INTENSIDAD DE TRANSVARIACION MARGINAL k -DIMENSIONAL

Consideramos primeramente las componentes de este concepto que son el momento de transvariación marginal k -dimensional y su valor máximo.

La suma de las intensidades de orden r , de todas las posibles transvariaciones marginales k -dimensionales, que en la realización del e.a. Λ pueden presentar la v.a. ξ y el parámetro A , con respecto al parámetro λ , considerando k componentes de la v.a. ξ , nos define el momento de transvariación marginal k -dimensional de orden r , de estas k componentes consideradas. Simbólicamente tenemos, para sus k primeras componentes

$$m_{r,n,k} = \int_{-\infty}^{A_1} dx_1 \int_{-\infty}^{A_2} dx_2 \dots \int_{-\infty}^{A_k} \Pi (A_i - x_i)^r f(x_1, x_2, \dots, x_k) dx_k \quad (5.7)$$

Para $r = 0$, tenemos definida la probabilidad de transvariación marginal k -dimensional.

Con $\beta_{r;n,k;A_i}$, simbolizamos el momento producto absoluto de orden k r de la distribución marginal con respecto al origen A_i . Luego

$$\begin{aligned} \beta_{r;n,k;A_i} = E(\Pi | x_i - A_i |^r) = m_{r;n,k} + \\ + \int_{A_1}^{\infty} dx_1 \int_{A_2}^{\infty} dx_2 \dots \int_{A_k}^{\infty} \Pi(x_i - A_i)^r f(x_1, x_2, \dots, x_k) dx_k \end{aligned} \quad (5.8)$$

Expresando el momento de transvariación marginal k -dimensional en función del parámetro $t = (t_1, t_2, \dots, t_n)$, tenemos

$$m_{r;n,k}(t) = \int_{-\infty}^{A_1+t_1} dx_1 \int_{-\infty}^{A_2+t_2} dx_2 \dots \int_{-\infty}^{A_k+t_k} \Pi(A_i + t_i - x_i)^r f(x_1, x_2, \dots, x_k) dx_k \quad (5.9)$$

cuyo máximo lo obtenemos para $t_i = \lambda_i - A_i$; $i = 1, 2, \dots, k$. Luego

$$m_{r;n,k;M} = \int_{-\infty}^{\lambda_1} dx_1 \int_{-\infty}^{\lambda_2} dx_2 \dots \int_{-\infty}^{\lambda_k} \Pi(\lambda_i - x_i)^r f(x_1, x_2, \dots, x_k) dx_k \quad (5.10)$$

La raíz r -ésima del cociente entre el momento de transvariación marginal k -dimensional de orden r y su valor máximo, nos define la intensidad de transvariación marginal k -dimensional de orden r . Simbolizándola con $I_{r;n,k}$ tenemos

$$I_{r;n,k} = \frac{(m_{r;n,k})^{\frac{1}{r}}}{(m_{r;n,k;M})^{\frac{1}{r}}} = \frac{m_{(r;n,k)}}{m_{(r;n,k;M)}} \quad (5.11)$$

Para $r = 1$, tenemos

$$I_{1;n,k} = \frac{m_{1;n,k}}{m_{1;n,k;M}} \quad (5.12)$$

Para $k = 1$, la (5.11) nos define la intensidad de transvariación marginal unidimensional de grado r , que coincide plenamente con la (2.17).

Las características de la intensidad de transvariación unidimensional dadas en 2.7, se pueden fácilmente hacer extensivas al presente caso general del cual constituye un caso particular.

Para $k = n$, tenemos

$$I_{r,n} = \frac{(m_{r;n})^{\frac{1}{r}}}{(m_{r;n;M})^{\frac{1}{r}}} = \frac{m_{(r;n)}}{m_{(r;n;M)}} \quad (5.13)$$

que nos define la intensidad de transvariación n -dimensional.

Si consideramos el espacio hilbertiano, estaríamos siempre en presencia de intensidades de transvariación marginales.

5.4. TRANSVARIACION MARGINAL ENTRE VARIABLES ALEATORIAS INDEPENDIENTES

Si las componentes de la v.a. ξ son v.a. independientes, la (5.3), (5.5) y (5.6), por definición de v.a. independientes, toman respectivamente, la siguiente expresión.

$$p_{n,k} = F(A_1, A_2, \dots, A_k) = \prod_{i=1}^k F_i(A_i) \quad (5.14)$$

$$p_{n,k;M} = F(\lambda_1, \lambda_2, \dots, \lambda_k) = \prod_{i=1}^k F_i(\lambda_i) \quad (5.15)$$

$$P_{n,k} = \frac{p_{n,k}}{p_{n,k;M}} = \prod_{i=1}^k \frac{F_i(A_i)}{F_i(\lambda_i)} = \prod_{i=1}^k P_{n;i} \quad (5.16)$$

donde con $P_{n;i}$ simbolizamos la probabilidad de transvariación marginal unidimensional de la v.a. ξ_i . Este resultado nos dice que, para el caso de v.a. independientes, la probabilidad de transvariación marginal k -dimensional es igual al producto de las probabilidades de transvariación de cada una de las k v.a. consideradas.

A partir de (5.7), (5.8), (5.10) y (5.11), tenemos respectivamente, por tratarse de *v.a.* independientes y de integrales definidas

$$m_{r;n,k} = \prod_{i=1}^k \int_{-\infty}^{A_i} (A_i - x_i)^r f_i(x_i) dx_i = \prod_{i=1}^k m_{r,i} \quad (5.17)$$

$$\beta_{r;n,k;A_i} = \prod_{i=1}^k E(|x_i - A_i|^r) = \quad (5.18)$$

$$= \prod_{i=1}^k \left[m_{r,i} + \int_{A_i}^{\infty} (x_i - A_i)^r f_i(x_i) dx_i \right]$$

$$m_{r;n,k;M} = \prod_{i=1}^k \int_{-\infty}^{\lambda_i} (\lambda_i - x_i)^r f_i(x_i) dx_i = \prod_{i=1}^k m_{r;M,i} \quad (5.19)$$

$$I_{r;n,k} = \prod_{i=1}^k \frac{\frac{1}{m_{r,i}}}{\frac{1}{m_{r;M,i}}} = \prod_{i=1}^k I_{r,i} \quad (5.20)$$

donde con $m_{r,i}$ simbolizamos el momento de transvariación marginal unidimensional de la *v.a.* ξ_i ; con $m_{r;M,i}$ su valor máximo y con $I_{r,i}$ su correspondiente intensidad de transvariación marginal unidimensional.

La (5.20) nos dice que también la intensidad de transvariación marginal, cuando las *v.a.* son independientes, es igual al producto de sus correspondientes intensidades de transvariación unidimensional.

Simbolizando con M_i la media aritmética de la *v.a.* ξ_i , tenemos para el momento producto absoluto de orden kr de la distribución marginal, con respecto a sus correspondientes medias aritméticas.

$$\beta_{r;n,k} = \prod_{i=1}^k E(|\xi_i - M_i|^r) = \prod_{i=1}^k \beta_{r,i}$$

donde con $\beta_{r,i}$ simbolizamos el momento absoluto de orden r de la v.a. $\xi_i - M_i$.

Si las ξ_i , para $i = 1, 2, \dots, n$; son simétricas, se verifica

$$m_{r,M,i} = \frac{1}{2} \beta_{r,i} \quad ; \quad (i = 1, 2, \dots, n)$$

5.5. MEDIDA SINTETICA DE LA TRANSVARIACIÓN EN R_n

En general, tanto la probabilidad como la intensidad de transvariación marginal k -dimensional, nos da una medida de la tipicidad de las diferencias entre los parámetros λ_i y A_i del conjunto de las k correspondientes v.a. unidimensionales consideradas. Si desde el punto de vista teórico el problema tiene una solución rigurosa, desde el punto de vista de las aplicaciones ofrece, con el crecer del número de dimensiones, dos serios inconvenientes, a saber:

1) necesita un creciente número de observaciones para reducir a un mínimo aceptable la creciente variabilidad debida al azar.

2) crece en complejidad el trabajo operatorio.

Para obviar estos inconvenientes y siguiendo a Gini (Ref. 13), definimos la v.a. reducida.

$$\tau = \tau_1 + \tau_2 + \dots + \tau_n = \sum_{i=1}^n \tau_i = \begin{cases} \sum \frac{\xi_i}{\Delta_i} \text{ (I)} \\ \sum \frac{\xi_i}{\sigma_i} \text{ (II)} \end{cases} \quad (5.21)$$

donde

$$\tau_i = \begin{cases} \frac{\xi_i}{\Delta_i} \text{ (I)} \\ \frac{\xi_i}{\sigma_i} \text{ (II)} \end{cases} \quad (5.21.a)$$

siendo

$$t = \sum_{i=1}^n t_i \quad (5.22)$$

donde

$$t_i = \begin{cases} \frac{x_i}{\Delta_i} & \text{(I)} \\ \frac{x_i}{\sigma_i} & \text{(II)} \end{cases} \quad (5.22.a)$$

es una particular determinación de la v.a. τ_i . Los parámetros Δ_i y σ_i nos definen, respectivamente, la diferencia media y el desvío medio cuadrático de la v.a. ξ_i . Podemos trabajar indistintamente con cualquiera de ellos; en adelante solo lo haremos con la diferencia media, salvo especificación en contrario.

Haciendo

$$(I) \begin{cases} A = \Sigma \frac{A_i}{\Delta_i} \\ \lambda = \Sigma \frac{\lambda_i}{\Delta_i} \end{cases} \quad (II) \begin{cases} A = \Sigma \frac{A_i}{\sigma_i} \\ \lambda = \Sigma \frac{\lambda_i}{\sigma_i} \end{cases} \quad (5.23)$$

y simbolizando respectivamente con $\tau_{(k)}$, $t_{(k)}$, $A_{(k)}$, y $\lambda_{(k)}$, a la v.a. τ , su particular determinación en el muestreo t y a los parámetros A y λ , cuando consideremos solamente k de las n componentes, tendremos, por definición, que la transvariabilidad marginal k dimensional, considerando sus k primeras componentes, es igual a la probabilidad de que en la realización del e.a. Λ , la v.a. $\tau_{(k)}$ asuma un valor $t_{(k)}$ tal que la diferencia

$$t_{(k)} - A_{(k)}$$

sea de signo opuesto al signo de la diferencia

$$\lambda_{(k)} - A_{(k)}$$

debiendo tomarse las componentes t_i , A_i y λ_i , para $i = 1, 2, \dots, k$, con el signo que tengan, cuando sea $\lambda_i > A_i$ y con signo opuesto, cuando sea $\lambda_i < A_i$. De la definición precedente, surgen en forma inmediata las expresiones analíticas de la transvariabilidad marginal, su valor máximo y de la probabilidad de transvariación marginal k dimensional.

Elas son, respectivamente para

$$\lambda_{(k)} > A_{(k)}$$

$$p_{n,(k)} = \int_{-\infty}^{A_{(k)}} g_k(t) dt \quad (5.24)$$

$$p_{n,(k);M} = \int_{-\infty}^{\lambda_{(k)}} g_k(t) dt \quad (5.25)$$

$$P_{n,(k)} = \frac{p_{n,(k)}}{p_{n,(k);M}} \quad (5.26)$$

donde con $g_k(t)$, simbolizamos la $f.$ de $d.$ de la v.a. $t_{(k)}$.

Para el momento de transvariación marginal k -dimensional de orden r , su valor máximo y para la intensidad de transvariación marginal k -dimensional de orden r , tenemos, respectivamente.

$$m_{r;n,(k)} = \int_{-\infty}^{A_{(k)}} (A_{(k)} - t_{(k)})^r g_k(t) dt \quad (5.27)$$

$$m_{r;n,(k);M} = \int_{-\infty}^{\lambda_{(k)}} (\lambda_{(k)} - t_{(k)})^r g_k(t) dt \quad (5.28)$$

$$I_{r;n,(k)} = \frac{(m_{r;n,(k)})^{\frac{1}{r}}}{(m_{r;n,(k);M})^{\frac{1}{r}}} \quad (5.29)$$

Dando a k valores desde 1 a $n-1$, tendremos, respectivamente, los correspondientes parámetros marginales de la transvariación desde 1 a $n-1$ dimensiones. Para $k=n$ tendremos definido los parámetros de la transvariación en el espacio R_n . Si consideramos el espacio hilbertiano, estaremos siempre en presencia de transvariaciones marginales. Para ello basta con suprimir la n en las

fórmulas precedentes, mientras que tendríamos, respectivamente para τ , t , A y λ ;

$$\tau = \sum_{i=1}^{\infty} \tau_i$$

$$t = \sum_{i=1}^{\infty} t_i$$

$$A = \sum_{i=1}^{\infty} \frac{A_i}{\Delta_i}$$

$$\lambda = \sum_{i=1}^{\infty} \frac{\lambda_i}{\Delta_i}$$

Al dividir por Δ_i o σ_i , convertimos a las variables aleatorias y a los parámetros, en v.a. y parámetros de dimensión nula lo que hace posible y da sentido a la suma, pues los vuelve comparables entre sí, por ser todos de la misma dimensión.

Todo lo estudiado en el capítulo 2, sobre la transvariación en R_1 , se extiende sin mayores variantes al presente parágrafo, donde la v.a. sintética $\tau_{(k)}$ puede ser tratada a tal efecto como una v.a. unidimensional.

5.6. TRANSVARIACIÓN MARGINAL EN LA HIPOTESIS DE DISTRIBUCIÓN NORMAL.

Primer Método. — Dado el e.a. Λ , al cual asociamos la v.a. ξ , cuya f. de d. es

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \sigma_2 \dots \sigma_n \sqrt{D}} e^{-\frac{1}{2D} \sum_{i,j} D_{i,j} \frac{(x_i - M_i)}{\sigma_i} \cdot \frac{(x_j - M_j)}{\sigma_j}} \quad (5.30)$$

donde

$$D = \begin{vmatrix} \rho_{11} & \rho_{12} & \dots & \rho_{1n} \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \rho_{n1} & \rho_{n2} & \dots & \rho_{nn} \end{vmatrix} \quad (5.30a)$$

es el determinante simétrico de todos los coeficientes lineales de correlación ρ_{ij} , por ser $\rho_{ij} = \rho_{ji}$. Los elementos de la diagonal principal son todos iguales a 1, por ser ρ_{ii} el coeficiente de correlación de dos v.a. idénticas. Con D_{ij} , simbolizamos el adjunto del elemento ρ_{ij} en el determinante D .

La f. de p.a. es

$$F_{(x_1, x_2, \dots, x_n)} = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \sigma_2 \dots \sigma_n \sqrt{D}} \int_{-\infty}^{x_1} dt_1 \int_{-\infty}^{x_2} dt_2 \dots \dots \int_{-\infty}^{x_n} e^{-\frac{1}{2D} \sum_{i,j} D_{i,j} \frac{(t_i - M_i)}{\sigma_i} \cdot \frac{(t_j - M_j)}{\sigma_j}} dt_n \quad (5.31)$$

Considerando k cualquiera de las n componentes de la v.a. n -dimensional, tendremos la correspondiente distribución marginal de probabilidad k -dimensional. Al respecto recordamos que cualquier distribución marginal de una distribución normal es también normal.

La f. de p.a. de la v.a. reducida es

$$\Phi_{(x_1, x_2, \dots, x_n)} = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{D}} \int_{-\infty}^{x_1} dt_1 \int_{-\infty}^{x_2} dt_2 \dots \int_{-\infty}^{x_n} e^{-\frac{1}{2D} \sum_{i,j} D_{i,j} t_i t_j} dt_n \quad (5.32)$$

Haciendo $\lambda_i = M_i$ para $i = 1, 2, \dots, n$, siendo M_i igual a la media aritmética y al valor mediano de la v.a. ξ_i , por tratarse de una distribución simétrica, tenemos para la probabilidad de transvariación marginal k -dimensional, considerando las k primeras v.a.

$$P_{n,k} = \frac{p_{n,k}}{p_{n,k;M}} = \frac{F(A_1, A_2, \dots, A_k)}{F(M_1, M_2, \dots, M_k)} = 2F(A_1, A_2, \dots, A_k) \quad (5.33)$$

por ser

$$F(M_1, M_2, \dots, M_k) = \frac{1}{2}$$

por definición de hiperplano mediano.

Para la intensidad de transvariación marginal k -dimensional de orden r , tenemos

$$I_{r,n,k} = \frac{(m_{r,n,k})^{\frac{1}{r}}}{(m_{r,n,k;M})^{\frac{1}{r}}} = \frac{(2 m_{r,n,k})^{\frac{1}{r}}}{(\beta_{r,n,k})^{\frac{1}{r}}} \quad (5.34)$$

donde $m_{r,n,k}$, lo obtenemos a partir de (5.7); la f. de d ., en muestra hipótesis, está dada por la distribución marginal de probabilidad, normalmente distribuida, obtenida de (5.31) y

$$\beta_{r,n,k} = E \left[\left| \prod_{i=1}^k (x_i - M_i) \right|^r \right] \quad (5.35)$$

que nos define el momento producto absoluto de orden kr de la distribución normal marginal k -dimensional considerada. En virtud de la simetría de la distribución, es

$$m_{r,n,k;M} = \frac{1}{2} \beta_{r,n,k} \quad (5.36)$$

Si las componentes $\xi_1, \xi_2, \dots, \xi_n$, son v.a. independientes se demuestra que $D = D_{ii} = 1$, mientras que $D_{ij} = 0$ para $i \neq j$. La f. de p.a. es

$$F(x_1, x_2, \dots, x_n) = F_1(x_1) F_2(x_2) \dots F_n(x_n) = \quad (5.37)$$

$$= \frac{1}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^n \frac{1}{\sigma_i} \int_{-\infty}^{x_i} e^{-\frac{(t_i - M_i)^2}{2\sigma_i^2}} dt_i$$

La probabilidad de transvariación marginal k -dimensional es

$$\begin{aligned} P_{n,k} &= 2 \prod_{i=1}^k F_i(A_i) = \frac{2}{(2\pi)^{\frac{n}{2}}} \prod_{i=1}^k \int_{-\infty}^{\frac{A_i - M_i}{\sigma_i}} e^{-\frac{t_i^2}{2}} dt_i = \\ &= 2 \prod_{i=1}^k \Phi\left(\frac{A_i - M_i}{\sigma_i}\right) = \prod_{i=1}^k P_{n;i} \end{aligned} \quad (5.38)$$

Para la intensidad de transvariación marginal k -dimensional de orden r , tenemos

$$I_{r,n,k} = \frac{\left[\prod_{i=1}^k \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{A_i} (A_i - x_i)^r e^{-\frac{(x_i - M_i)^2}{2\sigma_i^2}} dx_i \right]^{\frac{1}{r}}}{\left[\prod_{i=1}^k \frac{1}{\sigma_i \sqrt{2\pi}} \int_{-\infty}^{M_i} (M_i - x_i)^r e^{-\frac{(x_i - M_i)^2}{2\sigma_i^2}} dx_i \right]^{\frac{1}{r}}} = 2^{\frac{1}{r}} \prod_{i=1}^k \frac{m_{r,i}^{\frac{1}{r}}}{\beta_{r,i}^{\frac{1}{r}}} = \prod_{i=1}^k I_{r,i} \quad (5.39)$$

donde $m_{r,i}$ lo obtenemos a partir de (2.49) y $\beta_{r,i}$ a partir de (2.51.a) o (2.51.b), según sea r par o impar.

Para $r = 1$, tenemos

$$I_{1,n,k} = \prod_{i=1}^k I_{1,i} = e^{-\frac{k}{2} \sum_{i=1}^k \frac{(A_i - M_i)^2}{\sigma_i^2}} + (2\pi)^{\frac{k}{2}} \prod_{i=1}^k \frac{A_i - M_i}{\sigma_i} \Phi\left(\frac{A_i - M_i}{\sigma_i}\right) \quad (5.40)$$

Segundo Método. Para el caso que las n v.a. unidimensionales sean normalmente independientes, el primer método nos permite arribar, en forma inmediata, con el auxilio de la tabla de la distribución normal, al valor de la probabilidad de transvariación marginal k -dimensional y al valor de la intensidad de transvariación marginal k -dimensional de primer orden. Para el caso de v.a. normalmente dependientes, el proceso operatorio se vuelve cada vez más prohibitivo con el crecer del número de dimensiones, inconveniente que se supera con toda facilidad, recurriendo a una transformación lineal, la que nos conducirá a la medida sintética de los parámetros de la transvariación marginal k -dimensional y de la transvariación n -dimensional.

En general, consideraremos la distribución normal en R_n , cuyas componentes se encuentran normalmente correlacionadas, deduciendo luego como caso particular la que responda a la hipótesis de independencia. Veamos antes cómo se distribuye una función lineal de v.a. normalmente distribuidas.

Tengamos las siguientes v.a. normalmente dependientes

$$\xi_1, \xi_2, \dots, \xi_n \quad (5.41)$$

con medias y varianzas respectivamente iguales a

$$M_1, M_2, \dots, M_n, \quad (5.41.a)$$

$$\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2 \quad (5.41.b)$$

siendo ρ_{ij} ($i, j = 1, 2, \dots, n$) el coeficiente de correlación entre ξ_i y ξ_j ; $\mathbf{D} = \{\rho_{i,j}\}$ la matriz cuadrada de correlación y $\mathbf{C} = \{y_{ij}\} = \{\rho_{ij} \sigma_i \sigma_j\}$ la matriz cuadrada de momentos, siendo ambas simétricas por ser $\rho_{ij} = \rho_{ji}$.

La f. de d. y la f. de p.a. correspondiente están dadas respectivamente, por (5.30) y (5.31) Su f.c. es

$$\Psi(u) = \Psi(u_1, u_2, \dots, u_n) = e^{i \sum_{j=1}^n M_j u_j - \frac{1}{2} \sum_{h,j} y_{hj} u_h u_j} \quad (5.42)$$

que con notación matricial nos da

$$\Psi(u) = e^{i \mathbf{M}' u - \frac{1}{2} u' \mathbf{C} u} \quad (5.42.a)$$

donde $\mathbf{M}' = \{M_1, M_2, \dots, M_n\}$ es el vector-fila, transpuesto del vector-columna $\mathbf{M} = (M_1, M_2, \dots, M_n)$ y u' , análogamente, el vector-fila transpuesto del vector-columna $u = (u_1, u_2, \dots, u_n)$.

Definimos la v.a.

$$\eta = \sum_{i=1}^n b_i \xi_i \quad (5.43)$$

que es función lineal de las n v.a. normalmente correlacionadas. Su media y varianza son, respectivamente

$$E(\eta) = \sum_{i=1}^n b_i E(\xi_i) = \sum_{i=1}^n b_i M_i = b M \quad (5.43.a)$$

$$b^2 \sigma^2 = D^2(\eta) = E(\eta - b M)^2 = \sum_{i=1}^n b_i^2 \sigma_i^2 + 2 \sum_{i < j} b_i b_j \rho_{ij} \sigma_i \sigma_j =$$

$$= \sum_{i=1}^n b_i^2 \sigma_i^2 + 2 \sum_{i < j} b_{ij} y_{ij}$$

Su f.c. es, por definición

$$\Psi_{\eta}(v) = E(e^{iv\eta}) = \Psi_{\Sigma \xi_i}(b_1 u_i, b_2 u_i, \dots, b_n u_i) \quad (5.44)$$

para $j = 1, 2, \dots, n$. Su expresión analítica se obtiene en forma inmediata, a partir de (5.42), lo que nos da, luego de hacer

$$u_1 = u_2 = \dots = u_n = v$$

$$\Psi_{\eta}(v) = e^{ibMv - \frac{1}{2} b^2 \sigma^2 v^2} \quad (5.44.a)$$

que es la f.c. de una *v.a.* normalmente distribuida, con media igual a (5.43.a) y con varianza igual a (5.43.b). Su f. de d. es

$$g(y) = \frac{1}{b \sigma \sqrt{2\pi}} e^{-\frac{(y-bM)^2}{2b^2 \sigma^2}} \quad (5.45)$$

Para el caso de que las *v.a.* definidas en (5.41) fuesen normalmente independientes, sabemos ya, por el teorema de composición de funciones características, que la *v.a.* η , definida en (5.43) se distribuye normalmente con media igual a bM , definidas en (5.43.a) y varianza

$$b^2 \sigma^2 = \sum_{i=1}^n b_i^2 \sigma_i^2 \quad (5.46)$$

resultado al que llegamos también, como un caso particular del precedente, anulando los coeficientes de correlación, lo que a su vez anula las covarianzas y_{ij} . Para el caso de distribuciones marginales, el resultado es completamente análogo, para lo que nos basta recordar que toda distribución marginal de una distribución normal es otra distribución normal.

Luego de las conclusiones precedentes, estudiamos la medida sintética de la transvariación marginal, en la hipótesis de distribución normal.

De acuerdo con la definición (II) de (5.21) y teniendo presente (5.43) y (5.44.a), la *v.a.*

$$\tau = \sum_{i=1}^n \tau_i = \sum_{i=1}^n \frac{\xi_i}{\sigma_i} \quad (5.47)$$

se distribuye normalmente, con media y varianza, respectivamente iguales a

$$m = \sum_{i=1}^n \frac{M_i}{\sigma_i} \quad (5.47.a)$$

$$\sigma^2 = n + 2 \sum_{i < j} \rho_{ij} = \sum_{i,j} \rho_{ij} \quad (i, j = 1, 2, \dots, n) \quad (5.47.b)$$

Empleando la notación dada en 5.5, tenemos respectivamente para la probabilidad de transvariación marginal k -dimensional y para la intensidad de transvariación marginal k -dimensional de orden r , con respecto a la media aritmética, las expresiones dadas en (5.26) y (5.29), luego de reemplazar $\lambda_{(k)}$ por $m_{(k)}$, siendo

$$m_{(k)} = \sum_{i=1}^k \pm \frac{M_i}{\sigma_i} \quad (5.48.a)$$

$$g_{(k)}(t) = \frac{1}{\sigma_{(k)} \sqrt{2\pi}} e^{-\frac{(t-m_{(k)})^2}{2\sigma_{(k)}^2}} \quad (5.48.b)$$

$$\sigma_{(k)}^2 = k + 2 \sum_{i < j} \rho_{i,j} = \sum_{i,j} \rho_{i,j} \quad (i, j = 1, 2, \dots, k) \quad (5.48.c)$$

donde los sumandos $\frac{M_i}{\sigma_i}$ de (5.48.a), de acuerdo con la convención adoptada en 5.5, se deben tomar con el signo que presentan o con signo contrario, según sea M_i mayor o menor de A_i .

Si las variables ξ_i son normalmente independientes, la (5.48.c) resulta

$$\sigma_{(k)}^2 = \sum_{i=1}^k \rho_{ii} = k$$

6. — TRANSVARIACION ENTRE DOS DISTRIBUCIONES DE n DIMENSIONES

6.1. — INTRODUCCION

Veamos a continuación el caso general de los parámetros marginales de la transvariación entre dos v.a. n -dimensionales, deduciendo luego su medida sintética.

Sea Λ_1 , un e.a. al cual asociamos la v.a. n -dimensional

$$\xi = (\xi_1, \xi_2, \dots, \xi_n) \quad (6.1)$$

cuya f. de p.a. es

$$F(x) = P(\xi \leq x) = F(x_1, x_2, \dots, x_n) \quad (6.1.a)$$

siendo

$$x = (x_1, x_2, \dots, x_n) \quad (6.1.b)$$

una particular determinación de la v.a. ξ .

Sea Λ_2 , otro e.a. al cual asociamos la v.a. n -dimensional

$$\eta = (\eta_1, \eta_2, \dots, \eta_n) \quad (6.2)$$

cuya f. de p.a. es

$$G(y) = P(\eta \leq y) = G(y_1, y_2, \dots, y_n) \quad (6.2.a)$$

siendo

$$y = (y_1, y_2, \dots, y_n) \quad (6.2.b)$$

una particular determinación de la v.a. η .

Sea Λ , el experimento combinado (Λ_1, Λ_2) consistente en una realización simultánea de los e.a. Λ_1 y Λ_2 , observándose los resultados de ambos. En consecuencia, al e.a. combinado Λ , está asociada una v.a. combinada

$$\omega = (\xi, \eta) = (\xi_1, \xi_2, \dots, \xi_n, \eta_1, \eta_2, \dots, \eta_n)$$

cuyas componentes son las $2n$ componentes de las v.a. ξ y η . El espacio de la v.a. combinada, es el espacio de $2n$ dimensiones, producto de los espacios n -dimensiones de ξ y η . Su f. de p.a. es

$$J(v) = P(\omega \leq v) = J(x, y) = J(x_1, x_2, \dots, x_n, y_1, y_2, \dots, y_n)$$

donde

$$v = (x, y)$$

es una particular determinación de la v.a. ω .

Introduciendo los parámetros, por ahora arbitrarios

$$\begin{cases} \lambda_x = (\lambda_{x_1}, \lambda_{x_2}, \dots, \lambda_{x_n}) \\ \lambda_y = (\lambda_{y_1}, \lambda_{y_2}, \dots, \lambda_{y_n}) \end{cases} \quad (6.3)$$

tenemos, por definición, una transvariación marginal de $2k$ dimensiones para $k < n$, en una realización determinada del e.a. combinado Λ , cuando el signo de la diferencia $x_i - y_i$, sea de signo opuesto al signo de la diferencia $\lambda_{x_i} - \lambda_{y_i}$, para k valores de i . El valor absoluto de la potencia

$$\pi (x_i - y_i)^r$$

extendido el producto a los k valores de i que se consideren, nos definirá su respectiva intensidad de orden r . Para $r = 1$ tenemos, al mismo tiempo que la intensidad de orden uno, la medida de dicha tranvariación.

Introduciendo la v.a.

$$\tau = (\tau_1, \tau_2, \dots, \tau_n) \quad (6.4.)$$

definiendo para $i = 1, 2, \dots, n$

$$\tau_i = \eta_i - \xi_i, \quad (6.4.a)$$

y siendo

$$t = (t_1, t_2, \dots, t_n) \quad (6.4.b)$$

una particular determinación de la v.a. τ , tenemos para su f. de p.a.

$$H(t_1, t_2, \dots, t_n) = \int_{-\infty}^{t_1} d u_1 \dots \int_{-\infty}^{t_n} d u_n \quad (6.5)$$

$$\int_{-\infty}^{\infty} d x_1 \dots \int_{-\infty}^{\infty} j(x_1, \dots, x_n, u_1 + x_1, \dots, u_n + x_n) d x_n$$

Si la f. de p.a. admite derivada de orden n , una vez con respecto a cada t_i , tendremos

$$\begin{aligned} h(t_1, t_2, \dots, t_n) &= \int_{-\infty}^{\infty} d x_1 \dots \\ &\dots \int_{-\infty}^{\infty} j(x_1, \dots, x_n, t_1 + x_1, \dots, t_n + x_n) d x_n \end{aligned} \quad (6.5a)$$

que nos define su correspondiente f. de d.

Si las v.a. (6.1.) y (6.2.) son independientes entre si, la (6.5), por definición de v.a. independientes, toma la forma

$$\begin{aligned} H(t) = H(t_1, \dots, t_n) &= \int_{-\infty}^{\infty} d x_1 \dots \\ &\dots \int_{-\infty}^{\infty} G(t_1 + x_1, \dots, t_n + x_n) f(x_1, \dots, x_n) d x_n \end{aligned} \quad (6.6)$$

Su función de densidad es, para el supuesto que exista la derivada de orden n , tomada una vez con respecto a cada t_i ,

$$\begin{aligned} h(t) = h(t_1, \dots, t_n) &= \int_{-\infty}^{\infty} d x_1 \dots \\ &\dots \int_{-\infty}^{\infty} g(t_1 + x_1, \dots, t_n + x_n) f(x_1, \dots, x_n) d x_n \end{aligned} \quad (6.6.a)$$

Definimos el parámetro

$$\lambda = (\lambda_1, \lambda_2, \dots, \lambda_n) \quad (6.7)$$

para el cual es

$$\lambda_i = \lambda_{x_i} - \lambda_{y_i} \quad (6.7.a)$$

con

$$M = (M_1, M_2, \dots, M_n) \quad (6.7.b)$$

simbolizamos la media aritmética de la (6.4), siendo

$$M_i = M_{\eta_i} - M_{\xi_i} \quad (6.7.c)$$

la media aritmética de la (6.4.a).

Con la introducción de la v.a. τ y el parámetro λ , volvemos nuevamente a la transvariación marginal k -dimensional. La definición de transvariación marginal k -dimensional, su intensidad de orden r y la medida de dicha transvariación, para esta nueva v.a. τ , se deducen en forma inmediata. En efecto, tendremos por definición, una transvariación marginal k -dimensional, cuando k -componentes de la v.a. n -dimensional τ , presenten simultáneamente valores del mismo signo que el signo de sus correspondientes k parámetros λ_i .

Su correspondiente intensidad de orden r , considerando sus k primeras componentes, es

$$\prod_{i=1}^k |t_i|^r$$

de donde

$$\prod_{i=1}^k |t_i|$$

nos da la medida de dicha intensidad, a la vez que la intensidad de orden uno.

6.2. — PROBABILIDAD DE TRANSVARIACIÓN MARGINAL

Definimos la transvariabilidad marginal k -dimensional entre la v.a. $\tau = (\tau_1, \dots, \tau_n)$ y el parámetro $\lambda = (\lambda_1, \dots, \lambda_n)$ a la probabilidad de que en una realización del e.a. combinado Λ , la v.a. τ asuma un valor t , tal que el signo de la diferencia $x_i - y_i = -t_i$ sea opuesto al signo de la diferencia $\lambda_{x_i} - \lambda_{y_i} = \lambda_i$, para k determinados valores de i .

Sin pérdida de generalidad supondremos que es $\lambda_{x_i} > \lambda_{y_i}$, para $i = 1, 2, \dots, n$.

Considerando las k primeras componentes de las v.a. ξ y η , tenemos, por definición de transvariabilidad marginal k -dimensional

$$\begin{aligned} \phi_{n,k} &= P(\xi_1 < \eta_1, \dots, \xi_k < \eta_k) = P(\tau_1 > 0, \dots, \tau_k > 0) = \\ &= \int_0^\infty dt_1 \dots \int_0^\infty h(t_1, \dots, t_k) dt_k \end{aligned} \quad (6.8)$$

Si el hiperplano de ecuación $t_1 + t_2 + \dots + t_k = 0$, es un hiperplano mediano, la transvariabilidad es igual a $\frac{1}{2}$.

Para su valor máximo tenemos

$$\phi_{n,k;M} = P(\tau_1 > -\lambda_1, \dots, \tau_k > -\lambda_k) \quad (6.9)$$

La probabilidad de transvariación marginal k -dimensional es

$$P_{n,k} = \frac{\phi_{n,k}}{\phi_{n,k;M}} \quad (6.10)$$

Para $k = 1$, tenemos la probabilidad de transvariación marginal unidimensional, cuyo valor coincide con (3.22), con lo que lo tratado en aquel capítulo, puede ser considerado como un caso particular de la teoría marginal de la transvariación entre dos distribuciones. Las características de la probabilidad de transvariación allí estudiadas, se pueden extender fácilmente al presente caso general.

6.3. — INTENSIDAD DE TRANSVARIACIÓN MARGINAL

Por definición de momento de transvariación marginal k -dimensional de orden r , considerando las k primeras componentes de la v.a. τ tenemos

$$m_{r,n,k} = \int_0^\infty dt_1 \dots \int_0^\infty \prod_{i=1}^k t_i^r h(t_1, \dots, t_k) dt_k \quad (6.11)$$

El momento de transvariación al igual que la probabilidad de transvariación marginal solo depende para cada función de distribución, del signo de λ_i , para $i = 1, 2, \dots, k$. Su valor no altera cualquiera sea la ubicación del vector $(\lambda_1, \lambda_2, \dots, \lambda_k)$ siempre que permanezca dentro de uno de los dos hiperplanos determinados por el hiperplano de ecuación $t_1 + t_2 + \dots + t_k = 0$

El máximo del momento de transvariación es

$$m_{r,n,k;M} = \int_{-\lambda_1}^{\infty} dt_1 \dots \int_{-\lambda_k}^{\infty} \prod_{i=1}^k t_i^r h(t_1, \dots, t_k) dt_k \quad (6.12)$$

El momento producto absoluto de orden $k \cdot r$, con respecto al origen λ_i , es

$$\beta_{r,n,k;\lambda_i} = E \left(\prod_{i=1}^k |\tau_i - \lambda_i|^r \right) \quad (6.13)$$

Por definición, tenemos, para la intensidad de transvariación marginal k -dimensional de orden r

$$I_{r,n,k} = \frac{(m_{r,n,k})^{\frac{1}{r}}}{(m_{r,n,k;M})^{\frac{1}{r}}} \quad (6.14)$$

Para $k = 1$, tenemos la intensidad de transvariación marginal unidimensional, cuya expresión analítica es igual a la (3.23).

Las características de la intensidad de transvariación estudiadas en 3.7, se pueden hacer extensivas, sin mayor dificultad, al presente caso de transvariación marginal k -dimensional, del cual, aquel puede ser considerado como un caso particular.

6.4. — ESPACIO DE TRANSVARIACIÓN MARGINAL

Una generalización del concepto de área de transvariación estudiado en el capítulo 4, al tratar la transvariación entre dos v.a. unidimensionales independientes, lo constituye el concepto de espacio de transvariación marginal.

A partir de las v.a. n -dimensionales por definidas (6.1) y (6.2), y en la hipótesis que ambas sean entre sí independientes, siendo en consecuencia la (6.6) su correspondiente f. de d., y

$$h(t_1, \dots, t_k) = \int_{-\infty}^{\infty} dx_1 \dots \quad (6.15)$$

$$\dots \int_{-\infty}^{\infty} f(x_1, \dots, x_k) g(t_1 + x_1, \dots, t_k + x_k) dx_k$$

su f. de d. marginal k -dimensional, definimos la función de frecuencia marginal $\delta(u_1, u_2, \dots, u_k)$, correspondiente a la v.a. $v = (v_1, v_2, \dots, v_k)$, asociada al e.a. Λ_3 , de la siguiente forma:

$$\delta(u_1, u_2, \dots, u_k) = N_1 f(u_1, u_2, \dots, u_k); \quad (6.16.a)$$

si es,

$$N_1 f(u_1, u_2, \dots, u_k) \leq N_2 g(u_1, u_2, \dots, u_k)$$

$$\delta(u_1, u_2, \dots, u_k) = N_2 g(u_1, u_2, \dots, u_k); \quad (6.16.b)$$

si es,

$$N_1 f(u_1, u_2, \dots, u_k) > N_2 g(u_1, u_2, \dots, u_k)$$

donde con N_1 y N_2 simbolizamos, respectivamente, el número de elementos de los colectivos descriptos por las v.a. ξ y η . La probabilidad de que en la realización del e.a. Λ_3 , la v.a. v asuma un valor u , perteneciente a la v.a. ξ , si se verifica la condición (6.16.a) o perteneciente a la v.a. η , si se verifica la (6.16.b), es

$$C_{n,k} = \frac{1}{N_1 + N_2} \int_{-\infty}^{\infty} du_1 \int_{-\infty}^{\infty} du_2 \dots \int_{-\infty}^{\infty} \delta(u_1, u_2, \dots, u_k) du_k \quad (6.17)$$

que definimos espacio de transvariación marginal k -dimensional.

Geométricamente, la probabilidad $C_{n,k}$ es igual a la medida del espacio común comprendido entre las dos distribuciones marginales y el hiperplano definido por sus respectivas k componentes.

El máximo de $C_{n,k}$ es no mayor de un medio. Es igual a un medio cuando $N_1 = N_2$ y las v.a. ξ y η son v.a. idénticas.

El cociente entre el espacio de transvariación marginal y su valor máximo en la hipótesis de *v.a.* idénticas, pertenecientes a colectivos igualmente numerosos, nos define el espacio relativo de transvariación marginal. En símbolos

$$H_{n,k} = \frac{2}{N_1 + N_2} \int_{-\infty}^{\infty} d u_1 \int_{-\infty}^{\infty} d u_2 \dots \int_{-\infty}^{\infty} \delta(u_1, u_2, \dots, u_k) d u_k \quad (6.18)$$

Para $k = 1$, el espacio relativo de transvariación marginal unidimensional es igual al área relativa de transvariación de la *v.a.* considerada, concepto del cual nos ocupamos en 4.4.

A semejanza de los observado para el área relativa de transvariación, se observa también que a una probabilidad de transvariación marginal nula, le corresponde un espacio relativo de transvariación igual a cero, mientras que la afirmación recíproca no es cierta.

Aquí también, el cálculo del espacio relativo de transvariación marginal se efectúa en forma inmediata, lo que representa desde el punto de vista de las aplicaciones, una indudable ventaja sobre la probabilidad y la intensidad de transvariación marginal, cuyo cálculo resulta prohibitivo con solo considerar tres atributos.

Las características del espacio relativo de transvariación marginal se pueden deducir como una extensión de las vistas en 4.4. para el área relativa de transvariación.

Para $k = n$, tenemos por definición, el espacio relativo de transvariación n -dimensional, que simbolizamos con H_n .

6.5. — VALOR DISCRIMINATIVO MARGINAL

Definimos con $z_{(1)} = (z_1, z_2, \dots, z_k)$ el valor discriminativo marginal k -dimensional de las *v.a.* ξ y η , independientes entre si, si para dicho valor de $x_{(k)} = (x_1, x_2, \dots, x_k)$ es mínimo el error $\varepsilon(x_1, x_2, \dots, x_k)$, definido por

$$\varepsilon(x_1, x_2, \dots, x_k) = \frac{n_1(x_1, x_2, \dots, x_k) + n_2(x_1, x_2, \dots, x_k)}{N_1 + N_2} \quad (6.19)$$

siendo

$$\begin{aligned} n_1(x_1, x_2, \dots, x_k) = \\ = N_1 \int_S f(u_1, u_2, \dots, u_k) du_1 du_2 \dots du_k \end{aligned} \quad (6.19.a)$$

$$\begin{aligned} n_2(x_1, x_2, \dots, x_k) = \\ = N_2 \int_{R_k - S} g(u_1, u_2, \dots, u_k) du_1, du_2, \dots, du_k \end{aligned} \quad (6.19.b)$$

donde S y $R_k - S$ son dos subespacios k -dimensionales, sobre los cuales se encuentra proyectada la masa unitaria originaria, definida en el espacio R_n . Estos subespacios están definidos en tal forma que si en (6.19.a) le corresponde a u_i asumir valores menores o iguales que x_i , su correspondiente intervalo de integración en (6.19.b) es (x_i, ∞) o viceversa, sin que en ningún momento se le imponga la condición que las k.v.a. deban asumir simultáneamente valores menores o iguales que x_i en (6.19.a) y mayores o iguales en (6.19.b) o viceversa, para $i = 1, 2, \dots, k$.

Para el caso particular de que el subespacio S , de k dimensiones, esté definido en tal forma que

$$\begin{cases} (\xi_1 \leq x_1, \xi_2 \leq x_2, \dots, \xi_k \leq x_k) \subset S \\ (\eta_1 \geq x_1, \eta_2 \geq x_2, \dots, \eta_k \geq x_k) \subset R_k - S \end{cases} \quad (6.20)$$

la (6.19.a) y la (6.19.b), toman respectivamente la forma

$$\begin{aligned} n_1(x_1, x_2, \dots, x_k) = N_1 \int_{-\infty}^{x_1} du_1 \int_{-\infty}^{x_2} du_2 \dots \\ \dots \int_{-\infty}^{x_k} f(u_1, u_2, \dots, u_k) du_k \end{aligned} \quad (6.21.a)$$

$$n_2(x_1, x_2, \dots, x_k) = N_2 \int_{x_1}^{\infty} d u_1 \int_{x_2}^{\infty} d u_2 \dots$$

$$\dots \int_{x_k}^{\infty} g(u_1, u_2, \dots, u_k) d u_k \quad (6.21.b)$$

El valor $z_{(k)} = (z_1, z_2, \dots, z_k)$ que haga mínima la función del error definida por (6.19), lo obtenemos haciendo jugar las condiciones de máximos y mínimos para funciones de varias variables. La condición necesaria pero no suficiente para que dicho mínimo exista, exige, que para dicho valor sean nulas sus derivadas parciales primeras con respecto a cada una de las k variables consideradas. En el supuesto que los subespacios S y $R_k - S$ satisfagan la definición (6.20), obtenemos

$$N_1 \int_{-\infty}^{x_1} d u_1 \dots \int_{-\infty}^{x_{i-1}} d u_{i-1} \int_{-\infty}^{x_{i+1}} d u_{i+1} \dots$$

$$\dots \int_{-\infty}^{x_k} f(u_1, \dots, u_{i-1}, x_i, u_{i+1}, \dots, u_k) d u_k -$$

$$- N_2 \int_{x_1}^{\infty} d u_1 \dots \int_{x_{i-1}}^{\infty} d u_{i-1} \int_{x_{i+1}}^{\infty} d u_{i+1} \dots$$

$$\dots \int_{x_k}^{\infty} g(u_1, \dots, u_{i-1}, x_i, u_{i+1}, \dots, u_k) d u_k = 0 \quad (6.22)$$

para $i = 1, 2, \dots, k$, con lo que obtenemos un sistema de k ecuaciones con k incógnitas. Si la solución del sistema resultante nos da mas de una raíz para aquella que nos dé un mínimo absoluto de (6.19), nos definirá el valor discriminativo marginal k -dimensional, definiéndonos este mínimo absoluto, el error del valor discriminativo marginal k -dimensional.

El error del valor discriminativo marginal es igual o mayor que el respectivo espacio relativo de transvariación marginal k -dimensional. Es igual cuando las dos distribuciones marginales de frecuencias presentan una sola intersección y mayor, cuando presentan más de una. Si no se intersecan en ninguna parte del espacio en que están definidas, ambos parámetros son nulos. Ello significa que para cada sistema de valores (u_1, u_2, \dots, u_k) , por lo menos una de las f. de d. $f(u_1, u_2, \dots, u_k)$ y $g(u_1, u_2, \dots, u_k)$ son nulas.

6.6. — TRANSVARIACIÓN MARGINAL ENTRE VARIABLES ALEATORIAS INDEPENDIENTES.

Si las n componentes de cada una de las v.a. n -dimensionales ξ y η , son independientes, se deduce en forma inmediata para la probabilidad de transvariación marginal k -dimensional, a partir de (6.10), teniendo presente (6.8), (6.9) y la definición de variables aleatorias independientes.

$$P_{n,k} = \prod_{i=1}^k \frac{P(\tau_i > 0)}{P(\tau_i > -\lambda_i)} = \prod_{i=1}^k P_{n,i} \quad (6.23)$$

Es decir que, la probabilidad de transvariación marginal k -dimensional es igual al producto de las probabilidades de transvariación marginal unidimensional entre cada una de las k.v.a. τ_i consideradas, cuando ellas son independientes entre sí.

Para la intensidad de transvariación marginal k -dimensional de orden r , tenemos a partir de (6.14) y teniendo presente (6.11), (6.12) y la definición de v.a. independientes

$$I_{r,n,k} = \prod_{i=1}^k \frac{\left[\int_0^{\infty} t^r h_i(t) dt \right]^{\frac{1}{r}}}{\left[\int_{-\lambda_i}^{\infty} t^r h_i(t) dt \right]^{\frac{1}{r}}} = \prod_{i=1}^k \frac{m_{r,i}^{\frac{1}{r}}}{m_{r,M,i}^{\frac{1}{r}}} = \prod_{i=1}^k I_{r,i} \quad (6.24)$$

Como vemos, también es igual al producto de las intensidades de transvariaciones marginales unidimensionales de orden r , de las k -v.a. consideradas.

Si $-\lambda_i$, es igual al valor mediano de τ_i , tenemos para la (6.23)

$$P_{n,k} = \prod_{i=1}^k P_{n,i} = 2^k \prod_{i=1}^k P(\tau_i > 0) \quad (6.25)$$

La (6.13), toma la forma

$$\beta_{r,n,k;\lambda_i} = \prod_{i=1}^k E(|\tau_i - \lambda_i|^r) = \prod_{i=1}^k \beta_{r;\lambda_i} \quad (6.26)$$

Si tomamos como origen la media aritmética definida en (6.7,c), para $i = 1, 2, \dots, n$, tenemos

$$\beta_{r,n,k;M_i} = \prod_{i=1}^k E(|\tau_i - M_i|^r) = \prod_{i=1}^k \beta_{r;M_i} \quad (6.26.a)$$

Siendo, con respecto al parámetro M_i

$$m_{r;M_i} = \frac{1}{2} \beta_{r;M_i} \quad (6.27)$$

si la v.a. τ_i es simétrica. Para esta hipótesis, la (6.24), toma la forma

$$I_{r,n,k} = \prod_{i=1}^k I_{r,i} = 2 \prod_{i=1}^k \frac{m_{r,i}^{\frac{1}{r}}}{\beta_{r,i}^{\frac{1}{r}}} \quad (6.28)$$

Las expresiones (6.23) y (6.24) presuponen solamente la independencia de las n componentes de cada una de las v.a. ξ y η , mientras que ξ_i y η_i , para $i = 1, 2, \dots, n$ pueden estar correlacionadas entre si. Si las v.a. ξ_i y η_i son entre sí independientes, para $i = 1, 2, \dots, n$, y simbolizando con N_{1i} y N_{2i} , respectivamente, el número de observaciones de los colectivos descriptos por las v.a. unidimensionales ξ_i y η_i , tenemos, para el espacio relativo de transvariación marginal k -dimensional definido en (6.18)

$$H_{n,k} = \prod_{i=1}^k \frac{2}{N_{1i} + N_{2i}} \int_{-\infty}^{\infty} \delta_i(u) du = \prod_{i=1}^k H_{1,i} \quad (6.29)$$

siendo

$$\left\{ \begin{array}{l} \delta_i(u) = N_{1i} f_i(u); \quad \text{si es } N_{1i} f_i(u) \leq N_{2i} g_i(u) \\ \delta_i(u) = N_{2i} g_i(u); \quad \text{si es } N_{1i} f_i(u) > N_{2i} g_i(u) \end{array} \right. \quad (6.29.a)$$

donde con $f_i(u)$ y $g_i(u)$ simbolizamos respectivamente las f. de d. de las v.a. unidimensionales ξ_i y η_i .

El factor $H_{1;i}$ de la (6.29) es igual y necesariamente debía serlo, en virtud de las hipótesis de las cuales partimos, a la (6.18) para cuando $k = 1$, considerando la misma componente unidimensional.

Definimos el error del valor discriminativo marginal k -dimensional para cuando las k v.a. consideradas en ξ y η son independientes, mediante el producto de los errores de los valores discriminativo marginales unidimensionales de las k v.a. consideradas. Para cada factor tendremos su correspondiente valor discriminativo marginal unidimensional z_i .

Conforme a la definición precedente y simbolizando con ϵ_k a dicho error, tenemos

$$\epsilon_k = \prod_{i=1}^k \epsilon(x_i) = \prod_{i=1}^k \frac{n_1(x_i) + n_2(x_i)}{N_1 + N_2} \quad (6.30)$$

donde $n_1(x_i)$ y $n_2(x_i)$, para $i = 1, 2, \dots, n$, se definen mediante (4.23) o (4.24) según corresponda, siendo sus respectivas f. de d. las correspondientes, respectivamente, a las variables ξ_i y η_i , ($i = 1, 2, \dots, k$).

Para cada factor de (6.30) obtenemos su valor discriminativo z_i , tomando la más favorable de las hipótesis (4.23) y (4.24), de acuerdo a lo ya visto en el parágrafo 4.5.

6.7. — MEDIDA SINTETICA DE LA TRANSVARIACIÓN MARGINAL.

Los parámetros de la transvariación marginal k -dimensional nos dan una medida de la tipicidad del conjunto de las diferencias entre los k pares de valores considerados en relación a cada una de sus respectivas v.a. Si estos valores son sus correspondientes medias, tendremos la medida de la tipicidad marginal del conjunto de las diferencias de los k pares de valores medios. Obser-

vamos ya la innegable ventaja que presentan los parámetros de la transvariación desde el punto de vista de las aplicaciones, cuando se toman con respecto a sus medias, pues los mismos condensan en una medida, simultaneamente, las diferencias entre las medias, las varianzas y las formas de las dos distribuciones sometidas a estudio.

Si estamos en presencia de dos distribuciones estadísticas de n dimensiones, no basta conocer los parámetros de la transvariación n dimensional, para tener una medida que nos sintetice la tipicidad de la diferencia entre las dos distribuciones, caracterizada a través de la diferencia entre sus valores medios; nos interesa conocer también la medida de la tipicidad del conjunto de k diferencias entre los valores medios correspondientes a k atributos, para valores de k desde 1 a $n - 1$, que vienen dados por los parámetros de la transvariación marginal k -dimensional. Supongamos, por vía de hipótesis, que queremos estimar el grado de eficiencia de dos políticas de desarrollo económico en dos países económicamente subdesarrollados, por ejemplo Argentina y Brasil, para referirnos a un problemas de gran actualidad a la vez que a dos países de ingentes riquezas potenciales y cuyo mas angustioso problema es la falta de capitales. El problema del desarrollo económico de las áreas subdesarrolladas, ha sido recientemente elevado al primer plano de la preocupación internacional, porque de él se espera, con mucho fundamento, la principal contribución a la tan esperada paz en el mundo, o por lo menos a la conquista de una sólida y armónica unión ideológica y armonía política entre los países no sometidos al comando de la U.R.S.S. Así lo han entendido los países de occidente y algunos de oriente como la India, escuchados y respetados en el mundo por la seriedad y autoridad de sus estadistas. Son excepciones a esta dirección, algunas potencias con veleidades imperialistas, fruto de una mentalidad colonial de siglos, que se resisten aún a escuchar la voz soberana de los pueblos del mundo, olvidando el significado ecuménico del tributo de sangre dado por los pueblos en nombre de la libertad y por sobre todo, no viven la hora del tiempo, no viven la realidad histórica del momento que se llama segunda mitad del siglo XX y no siglo XVI, XVII, XVIII o XIX. La importancia asignada al desarrollo económico lo confirma la mas alta autoridad internacional, como lo es la U.N., a través de la

consideración que le otorga a su estudio en los organismos técnicos que de ella dependen, como la C.E.P.A.L. (Comisión Económica Para América Latina).

Con lo precedentemente expuesto, no pretendemos embarcarnos con la insostenible tesis exclusiva del materialismo histórico o materialismo dialéctico, ni aún en su forma más amplia y menos materialista, llamada interpretación económica de la historia. Observamos ya que hasta en el dominio de las así llamadas ciencias exactas, ha sido completamente superado el esquema clásico de la causalidad, con cuanta mayor razón lo podemos considerar en el dominio de las ciencias sociales. Pero tampoco debemos negarle al César lo que es del César, por ello afirmamos que la interpretación económica de la historia es una de las componentes de primera magnitud en la explicación de la realidad social y en la consideración de su devenir y por ello señalamos la trascendencia histórica de las políticas de desarrollo económico apoyadas de buena fe por las naciones económicamente desarrolladas.

Volviendo a nuestro ejemplo supongamos que podemos explicar la realidad económica de Argentina y Brasil a través de dos distribuciones n -dimensionales, cuyas n variables podrían ser el potencial vida, el ingreso, el consumo, etc. Si conocemos dichas distribuciones en el momento t , para dicho momento del tiempo y con respecto a las medias de cada componente, podemos calcular los parámetros de la transvariación marginal. Estos nos darán una medida de la tipicidad del conjunto de las diferencias entre las medias, que nos permitirá una interpretación cuantitativa de la diferencia entre la realidad económica de ambos países en relación al grupo de variables consideradas.

Si a partir del momento t , ha actuado sobre cada uno de estos países una política de desarrollo económico durante un intervalo h de tiempo, conociendo las distribuciones respectivas en el momento $t + h$, podemos, a través de los parámetros de la transvariación marginal, estimar la tipicidad del conjunto de las diferencias entre las dos economías en el momento $t + h$ y considerando para cada país las distribuciones en el momento t y $t + h$, estimar la eficiencia de cada política de desarrollo.

Una primera aproximación al estudio de la realidad económica argentina-brasileña, la obtenemos admitiendo la presenta-

ción causal de los fenómenos en dicho espacio finito y suponiendo un sistema económico cerrado. Dando por cumplidas estas hipótesis, se puede determinar unívocamente la realidad económica en un momento $t + h$, conociendo dicha realidad en el momento t .

Una segunda aproximación y siempre dentro de los dominios de la causalidad, fecundamente superada por la concepción de la dependencia estocástica, la obtenemos haciendo uso de los aportes debidos a la teoría de la relatividad ⁽⁹⁾, que vuelve innecesaria la exigencia del sistema cerrado, exigiendo en cambio la existencia de leyes de acción próxima y un límite finito c de la velocidad de propagación de todas las fuerzas económicas. Como consecuencia, la formulación del determinismo económico implica la determinación unívoca de la realidad económica en el punto tempo-espacial $e_1 t_1$, conociendo la realidad económica en todos los puntos tempo-espaciales $e t$, correspondientes a los momentos de t anteriores a t_1 y a puntos e del espacio, tal que la distancia

$$|e - e_1| < c (t_1 - t)$$

Una tercera aproximación obtenida con los recursos de la teoría matemática de la probabilidad y por lo tanto superando el limitado marco de la causalidad o determinismo para penetrar en los amplios dominios de la dependencia estocástica, la conseguimos con el concurso de la Teoría de la Transvariación, en la forma precedentemente señalada.

El conjunto de los parámetros de la transvariación marginal de cada dimensión, simultáneamente con los parámetros de la transvariación n -dimensional, nos dará un sistema de medidas de gran eficacia, para interpretar la tipicidad de las diferencias entre las medias de cualquier grupo de variables y de cualquier número de dimensiones, con lo que no faltará nada para agotar su estudio cuantitativo.

(9) Para un meditado estudio de estos temas en los dominios de la Física y para una informada crítica al determinismo, nos remitimos a los trabajos de P. JORDAN: *La mecánica cuantista y los problemas fundamentales de la biología y de la psicología*, en «Revista de Occidente», Madrid, año X, N° CXIV, Dic. 1932 y A. EDDINGTON: *La decadencia del determinismo*, también «R. de Occidente», Setiembre 1932.

Queda en pié los problemas que plantea el tamaño de la muestra que pierde representatividad si no crece, con el crecer del número de dimensiones consideradas y en muchos casos no está en muestras manos la posibilidad de aumentar el tamaño de la muestra. Otro problema, prácticamente también insuperable para casi todos los parámetros de la transvariación, está dado por las dificultades prohibitivas del trabajo operatorio. Por tales razones estudiaremos la medidas sintéticas de la transvariación en forma análoga a la vista en el capítulo precedente.

Considerando las v.a. n -dimensionales, definidas en (6.1) y (6.2) y el sistema de parámetros definido en (6.3), introducimos respectivamente las siguientes v.a. reducidas n -dimensionales

$$\tau = \sum_{i=1}^n \tau_i = \sum_{i=1}^n \frac{\xi_i}{\Delta_{i,i}} \quad (6.31.a)$$

$$\nu = \sum_{i=1}^n \nu_i = \sum_{i=1}^n \frac{\eta_i}{\Delta_{i,i}} \quad (6.31.b)$$

donde con $\Delta_{i,i}$, simbolizamos la diferencia media de todas las posibles diferencias que se pueden establecer entre las v.a. ξ_i y η_i .

Considerando las k primeras componentes de ξ y η , definimos las v.a. reducidas marginales k -dimensionales.

$$\tau_{(k)} = \sum_{i=1}^k \tau_i = \sum_{i=1}^k \frac{\xi_i}{\Delta_{i,i}} \quad (6.32.a)$$

$$\nu_{(k)} = \sum_{i=1}^k \nu_i = \sum_{i=1}^k \frac{\eta_i}{\Delta_{i,i}} \quad (6.32.b)$$

Con

$$t_{(k)} = \sum_{i=1}^k t_i = \sum_{i=1}^k \frac{x_i}{\Delta_{i,i}} \quad (6.33.a)$$

$$u_{(k)} = \sum_{i=1}^k u_i = \sum_{i=1}^k \frac{y_i}{\Delta_{i,i}} \quad (6.33.b)$$

definimos una particular determinación, respectivamente, de la (6.32.) y (6.32.a), que se calculan observando la siguiente regla :

Para la componente t_i ($i = 1, 2, \dots, n$), sus valores reducidos serán tomados con el signo que resulten, si λ_i definido en (6.7.a), es positivo y con signo opuesto, si es negativo y recíprocamente, los correspondientes valores reducidos de u_i , serán tomados con signo opuesto o con el mismo signo, según que λ_i sea positivo o negativo.

Las v.a. $\tau_{(k)}$ y $v_{(k)}$ pueden ser consideradas como variables unidimensionales, cuyas particulares determinaciones están dadas por $t_{(k)}$ y $u_{(k)}$, calculadas de acuerdo a la regla precedente. Los parámetros de la transvariación entre estas v.a., nos definen los correspondientes parámetros marginales k -dimensionales. Si se encuentran correlacionadas, es de aplicación todo lo visto en el capítulo 3 y si en cambio ellas son independiente, es de aplicación todo lo visto en el capítulo 4.

Evidentemente para $k = 1$, obtendremos los mismos resultados sea que trabajemos con las v.a. reducidas o con sus correspondientes v.a. originales. En general, para dos distribuciones n -dimensionales, tenemos $\binom{n}{k}$ combinaciones posibles de grupos de v.a. de k dimensiones para cada uno de los cuales calculamos los parámetros marginales k -dimensionales de la transvariación. En resumen, para cada parámetro de la transvariación, tendremos $2^n - 1$ medidas, considerando las transvariaciones marginales de cada una de las dimensiones posibles y la transvariación n -dimensional.

6.8. — TRANSVARIACIÓN ENTRE DOS DISTRIBUCIONES NORMALES EN R_n

Dadas las v.a. n -dimensionales ξ y η definidas en (6.1.) y en (6.2), cada una de las cuales se encuentra normalmente distribuida, sus respectivas f. de d., f. de p.a. y f.c., toman la forma ya dada en (5.30), (5.31) y (5.42) para la v.a. ξ .

a) *Primer método*: Introducimos la v.a. τ en la forma ya vista en (6.4) y (6.4.a), con su particular determinación dada por (6.4.b) y expresamos mediante (6.7.c) la media de la v.a. τ_i . La f. de p.a. de la v.a. τ , se distribuye normalmente en el espacio

R_n , con media y varianza de su componente τ_i ($i = 1, 2, \dots, n$), respectivamente igual a

$$M_i = M_{\eta_i} - M_{\xi_i} \quad (6.34.a)$$

$$\sigma_i^2 = \sigma_{\eta_i}^2 + \sigma_{\xi_i}^2 - 2 \rho_i \sigma_{\eta_i} \sigma_{\xi_i} \quad (6.34.b)$$

donde con ρ_i simbolizamos el coeficiente de correlación entre las v.a. ξ_i y η_i , admitiendo la hipótesis que ellas se encuentran entre si correlacionadas. Podemos suponer que ξ_i y η_j , están también normalmente correlacionadas, para $i \neq j$, pero por tratarse de una hipótesis apartada de la realidad, las supondremos independientes.

En consecuencia

$$E(\xi_i - M_{\xi_i})(\eta_j - M_{\eta_j}) = \begin{cases} 0 & \text{para } i \neq j \\ \rho_i \sigma_{\xi_i} \sigma_{\eta_i} & \text{para } i = j \end{cases} \quad (6.34.c)$$

con lo que tenemos para τ_i y τ_j ($i, j = 1, 2, \dots, n$)

$$E(\tau_i - M_i)(\tau_j - M_j) = \rho_{\xi_i \xi_j} \sigma_{\xi_i} \sigma_{\xi_j} + \rho_{\eta_i \eta_j} \sigma_{\eta_i} \sigma_{\eta_j} = \rho_{ij} \sigma_i \sigma_j \quad (6.35)$$

Simbolizando con

$$\mathbf{B} = \{\rho_{ij}\} \quad (i, j = 1, 2, \dots, n) \quad (6.36.a)$$

la matriz de correlación de la v.a. τ , siendo

$$B = |\mathbf{B}| = |\rho_{ij}| \quad (6.36.b)$$

el determinante correspondiente a dicha matriz de correlación. Con B_{ij} , simbolizamos el adjunto de B .

De acuerdo con la notación precedente, la expresión analítica de la f. de d. de la v.a. τ , es:

$$h(t_1, t_2, \dots, t_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \sigma_2 \dots \sigma_n \sqrt{B}} e^{-\frac{1}{2B} \sum_{i,j} B_{ij} \frac{(t_i - M_i)}{\sigma_i} \frac{(t_j - M_j)}{\sigma_j}} \quad (6.37)$$

Su f. de d. marginal k -dimensional es :

$$h(t_1, t_2, \dots, t_k) = \frac{1}{(2\pi)^{\frac{n}{2}} \sigma_1 \sigma_2 \dots \sigma_k \sqrt{B_{(k)}}} e^{-\frac{1}{2B_{(k)}} \sum_{i,j} B_{(k)ij} \frac{(t_i - M_i)}{\sigma_i} \frac{(t_j - M_j)}{\sigma_j}} \quad (6.37.a)$$

donde con $B_{(k)}$ y $B_{(k)ij}$ simbolizamos el determinante de correlación de las k v.a. consideradas y su adjunto.

Ya disponemos de todos los elementos necesarios para calcular los parámetros de la transvariación marginal k -dimensional en la hipótesis de dos distribuciones normalmente correlacionadas. En efecto, la probabilidad de transvariación marginal k -dimensional la obtenemos llevando a (6.10) la (6.37.a). Si reemplazamos λ_i por M_i , la tendremos calculada con respecto a sus k -medias. Procediendo en forma análoga con la (6.14), llegamos a la intensidad de transvariación marginal k -dimensional de orden r .

Si admitimos la hipótesis que las dos distribuciones son entre sí independientes, haya o no dependencia estocástica entre las v.a. ξ_i y η_i ($i = 1, 2, \dots, n$), podemos deducir además de la probabilidad e intensidad de transvariación ya consideradas, el espacio de transvariación marginal, el valor discriminativo marginal y el error de dicho valor discriminativo, sin más que reemplazar en sus respectivas formulas vistas en 6.4 y en 6.5, la f. de d. marginal de la distribución normal de las v.a. ξ y η .

Si las v.a. τ_i y τ_j son entre si independientes, para cualquier combinación de i y de j , sean o no independientes las v.a. ξ_i y η_i , la (6.37.a) resulta

$$h(t_1, t_2, \dots, t_k) = \prod_{i=1}^k h_i(t) = \prod_{i=1}^k \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(t_i - M_i)^2}{2\sigma_i^2}} \quad (6.38)$$

con lo que tenemos, trabajando con respecto a las medias M_i , para la probabilidad de transvariación marginal k -dimensional, en virtud de (6.23) y (3.46)

$$P_{n,k} = \prod_{i=1}^k P_{n,i} = 2^k \prod_{i=1}^k \Phi(d_i) \quad (6.39)$$

siendo $\Phi_i(t)$ la f. de p.a. de la variable normal (0,1) y

$$d_i = \frac{M_i}{\sigma_i} \quad (6.39.a)$$

en la hipótesis

$$M_i = M_{\eta_i} - M_{\xi_i} < 0 \quad (6.39.b)$$

Si en cambio es $M_i > 0$ solo debemos reemplazar d_i por $-d_i$ en (6.39). La varianza σ_i^2 entre las v.a. ξ_i y η_i ($i = 1, 2, \dots, n$) está dada por la (6.34.b), si están normalmente correlacionadas y por

$$\sigma_i^2 = \sigma_{\eta_i}^2 + \sigma_{\xi_i}^2 \quad (6.40)$$

si son normalmente independientes, por ser nula en esta hipótesis la (6.34.c) para cualquier combinación de las v.a. ξ_i y η_j .

Para la intensidad de transvariación marginal k dimensional tenemos, a partir de (6.24) y (3.47) en la hipótesis (6.39.b)

$$I_{r,n,k} = \prod_{i=1}^k I_{r,i} = \prod_{i=1}^k \frac{\left[M_i^r \Phi(d_i) + \frac{1}{\sqrt{2\pi}} \sum_{s=1}^r \binom{r}{s} M_i^{r-s} \sigma_i^{\frac{s}{2}} G(s, d_i) \right]^{\frac{1}{r}}}{\sigma_i \sqrt{2} \left[\frac{1}{2\sqrt{\pi}} \Gamma\left(\frac{r+1}{2}\right) \right]^{\frac{1}{r}}} \quad (6.41)$$

donde d_i está definido por (6.39.a) y σ_i^2 por (6.34.b) o (6.40) según que las v.a. ξ_i y η_i ($i = 1, 2, \dots, n$) están correlacionadas o sean normalmente independientes. Si las v.a. ξ_i y η_i ($i = 1, 2, \dots, n$) son independientes la (6.29) nos da el espacio de transvariación marginal k -dimensional, donde

$$H_{1,i} = \frac{2}{N_{1,i} + N_{2,i}} \int_{-\infty}^{\infty} \delta_i(u) du \quad (6.42)$$

Siendo

$$\delta_i(u) = \frac{N_{1,i}}{\sigma_{\xi_i} \sqrt{2\pi}} e^{-\frac{(u-M_{\xi_i})^2}{2\sigma_{\xi_i}^2}} \quad \text{si es} \quad \frac{N_{1,i}}{\sigma_{\xi_i} \sqrt{2\pi}} e^{-\frac{(u-M_{\xi_i})^2}{2\sigma_{\xi_i}^2}} \leq \frac{N_{2,i}}{\sigma_{\eta_i} \sqrt{2\pi}} e^{-\frac{(u-M_{\eta_i})^2}{2\sigma_{\eta_i}^2}} \quad (6.42.a)$$

$$\delta_i(u) = \frac{N_{2,i}}{\sigma_{\eta_i} \sqrt{2\pi}} e^{-\frac{(u-M_{\eta_i})^2}{2\sigma_{\eta_i}^2}} \quad \text{si es} \quad \frac{N_{1,i}}{\sigma_{\xi_i} \sqrt{2\pi}} e^{-\frac{(u-M_{\xi_i})^2}{2\sigma_{\xi_i}^2}} > \frac{N_{2,i}}{\sigma_{\eta_i} \sqrt{2\pi}} e^{-\frac{(u-M_{\eta_i})^2}{2\sigma_{\eta_i}^2}}$$

Análogamente, admitiendo la hipótesis complementaria precedente, tenemos, a partir de la (6.30), en combinación con la (4.23) o (4.24), según corresponda, el valor discriminativo marginal k -dimensional y su error, donde sus respectivas f. de d. son las correspondientes a las v.a. ξ_i , normal $(M_{\xi_i}, \sigma_{\xi_i}^2)$ y η_i normal $(M_{\eta_i}, \sigma_{\eta_i}^2)$ para $i = 1, 2, \dots, k$.

b) *Segundo metodo*: Veamos a continuación como se resume todo lo precedente, trabajando con la medida sintética de la transvariación marginal, estudiada en el parágrafo anterior.

Siendo ξ y η dos v.a. n -dimensionales, normalmente distribuidas en R_n , introducimos las v.a. τ y ν en la forma definida en (6.31.a) y (6.31.b). Estas nuevas variables, por lo demostrado en el último parágrafo del capítulo anterior, se distribuyen normalmente con medias

$$M_1 = E(\tau) = \sum_{i=1}^n E(\tau_i) = \sum_{i=1}^n \frac{M_{\xi_i}}{\Delta_{ii}} = \sum_{i=1}^n M_{\tau_i} \quad (6.43.a)$$

$$M_2 = E(\nu) = \sum_{i=1}^n E(\nu_i) = \sum_{i=1}^n \frac{M_{\eta_i}}{\Delta_{ii}} = \sum_{i=1}^n M_{\nu_i} \quad (6.43.b)$$

y varianzas

$$\sigma_1^2 = D^2(\tau) = \sum_{i=1}^n \frac{\sigma_{\xi_i}^2}{\Delta_{ii}^2} + 2 \sum_{i < j} \frac{\rho_{ij} \sigma_{\xi_i} \sigma_{\xi_j}}{\Delta_{ii} \Delta_{jj}} \quad (6.44.a)$$

$$\sigma_2^2 = D^2(\nu) = \sum_{i=1}^n \frac{\sigma_{\eta_i}^2}{\Delta_{ii}^2} + 2 \sum_{i < j} \frac{\rho'_{ij} \sigma_{\eta_i} \sigma_{\eta_j}}{\Delta_{ii} \Delta_{jj}} \quad (6.44.b)$$

A partir de estas fórmulas obtenemos en forma inmediata las medias y varianzas de las distribuciones marginales k -dimensionales de las variables τ y ν , las que simbolizamos, respectivamente, con $M_{1(k)}$, $M_{2(k)}$, $\sigma_{1(k)}^2$ y $\sigma_{2(k)}^2$.

En (6.43.a) y (6.43.b), para cada componente de τ y ν , y consiguientemente para sus respectivas variables reducidas marginales k -dimensionales, se tomará con su signo la media mayor y con signo opuesto la media menor, haciendo corresponder igual signo al sumando correspondiente de τ y ν . Esta convención se tendrá también presente en (6.44.a) y (6.44.b), para determi-

nar el signo de los dobles productos en el último término de cada uno de ellos. Si las n componentes de cada una de las v.a. ξ y η son entre si independientes, estos dobles productos son nulos.

Las variables reducidas τ y ν , observando la convención precedente, pueden ser tratadas como v.a. unidimensionales y por lo tanto es de aplicación todo lo visto en 3.9 si ellas se encuentran entre sí correlacionadas y lo visto en 4.6, si son independientes.

Considerando las k primeras componentes de τ y ν y suponiendo, sin pérdida de generalidad, que es $M_{1(k)} > M_{2(k)}$ la probabilidad de transvariación marginal k -dimensional entre las v.a. reducidas $\tau_{(k)}$ $\nu_{(k)}$ está dada por (3.46) o (3.56) y la intensidad de transvariación marginal k -dimensional de orden r , por (3.47) o (3.57), donde

$$d = M_{1(k)} - M_{2(k)} > 0 \quad (6.45.a)$$

$$d_1 = \frac{d}{\sigma_{(k)}} \quad (6.45.b)$$

siendo

$$\sigma_{(k)}^2 = \sigma_{1(k)}^2 + \sigma_{2(k)}^2 - 2 \rho_{(k)} \sigma_{1(k)} \sigma_{2(k)} \quad (6.46.a)$$

si las v.a. $\tau_{(k)}$ y $\nu_{(k)}$ están normalmente correlacionadas y

$$\sigma_{(k)}^2 = \sigma_{1(k)}^2 + \sigma_{2(k)}^2 \quad (6.46.b)$$

si son entre sí independientes.

Para esta hipótesis de independencia, deducimos también el espacio relativo de transvariación marginal k -dimensional, el valor discriminativo y el error del valor discriminativo marginal k -dimensional, a partir de los resultados obtenidos en 4.6.

La f. de d. de las v.a. $\tau_{(k)}$ y $\nu_{(k)}$, si se encuentran correlacionadas, es

$$f(t_{(k)}, u_{(k)}) = \frac{1}{2 \pi \sigma_{1(k)} \sigma_{2(k)} \sqrt{1 - \rho_{(k)}^2}} \cdot \quad (6.47)$$

$$\cdot e^{-\frac{1}{2(1-\rho_{(k)}^2)} \left[\frac{(t_{(k)} - M_{1(k)})^2}{\sigma_{1(k)}^2} - 2\rho_{(k)} \frac{t_{(k)} - M_{1(k)}}{\sigma_{1(k)}} \frac{u_{(k)} - M_{2(k)}}{\sigma_{2(k)}} + \frac{(u_{(k)} - M_{2(k)})^2}{\sigma_{2(k)}^2} \right]}$$

Para $\rho_{(k)} = 0$, tenemos

$$f(t_{(k)}, u_{(k)}) = f(t_{(k)}) g(u_{(k)}) = \frac{1}{\sigma_{1(k)} \sqrt{2\pi}} e^{-\frac{(t_{(k)} - M_{1(k)})^2}{2\sigma_{1(k)}^2}} \cdot \frac{1}{\sigma_{2(k)} \sqrt{2\pi}} e^{-\frac{(u_{(k)} - M_{2(k)})^2}{2\sigma_{2(k)}^2}}. \quad (6.48)$$

que es la f. de d. de las v.a. $\tau_{(k)}$ y $\nu_{(k)}$, en la hipótesis que estas sean normalmente independientes.

Siempre en el supuesto $M_{1(k)} > M_{2(k)}$, si introducimos la nueva v.a.

$$\theta_{(k)} = \nu_{(k)} - \tau_{(k)} \quad (6.49)$$

demostramos en 5.6, en forma general, a partir de la función característica, que $\theta_{(k)}$ también se distribuye normalmente, con media

$$M_{(k)} = M_{2(k)} - M_{1(k)} = -d \quad (6.50)$$

y varianza dada por (6.46.a), si las v.a. $\nu_{(k)}$ y $\tau_{(k)}$ están correlacionadas y con varianza dada por (6.46.b), si son normalmente independientes. Su f. de d. es

$$h_{(\nu_{(k)})} = \frac{1}{\sigma_{(k)} \sqrt{2\pi}} e^{-\frac{(\nu_{(k)} - M_{(k)})^2}{2\sigma_{(k)}^2}} \quad (6.51)$$

donde $\nu_{(k)}$ es una particular determinación de la v.a. $\theta_{(k)}$.

7. — TRANSVARIACION ENTRE VARIAS DISTRIBUCIONES

7.1. — INTRODUCCIÓN

Sea Λ_i ($i = 1, 2, \dots, s$) un e.a. al cual se encuentra asociada la v.a. n -dimensional

$$\xi_i = (\xi_{i,1}, \xi_{i,2}, \dots, \xi_{i,n}) \quad (7.1)$$

definida en el espacio R_n , siendo

$$x_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n}) \quad (7.1.a)$$

una particular determinación de ella. Su f. de p.a. es

$$P(\xi_i \leq x_i) = F_i(x_i) = F(x_{i,1}, \dots, x_{i,n}) \quad (7.2)$$

Su f. de d. es

$$f_i(x_i) = f_i(x_{i,1}, \dots, x_{i,n})$$

Si la distribución es de tipo continuo, la f. de p.a. es continua en todo punto del espacio R_n , verificándose

$$f_i(x_{i,1}, \dots, x_{i,n}) = \frac{\delta^n F_i}{\delta x_{i,1}, \dots, \delta x_{i,n}}$$

la que es también continua en cualquier punto, excepto, a lo más en un conjunto finito de puntos que pertenecen a un número finito de hipersuperficies de R_n .

Simbolizando con N_i el número de observaciones del colectivo descrito por la v.a. ξ_i , su función de frecuencia es

$$N_i f_i(x_i) \quad (7.3)$$

Sea $\Lambda_{i,j} = (\Lambda_i, \Lambda_j)$; $i \neq j$, un e.a. combinado al cual asociamos la v.a. combinada $\theta_{i,j} = (\xi_i, \xi_j)$, definida en el espacio R_{2n} producto de los espacios n -dimensionales de las v.a. ξ_i y ξ_j . Su f. de p.a. es

$$P(\theta_{ij} \leq y_{ij}) = P(\xi_i \leq x_i, \xi_j \leq x_j) = F_{i,j}(x_i, x_j) \quad (7.4)$$

siendo

$$y_{ij} = (x_i, x_j)$$

una particular determinación de la v.a. θ_{ij} .

Si las v.a. ξ_i y ξ_j son entre si independientes resulta

$$F_{ij}(x_i, x_j) = F_i(x_i) F_j(x_j)$$

A cada v.a. ξ_i , ($i = 1, 2, \dots, s$), asociamos un parámetro real n -dimensional, por ahora arbitrario

$$\lambda'_i = (\lambda_{i1}, \dots, \lambda_{in}) \quad (7.5)$$

Introducimos la v.a.

$$\tau_{ij} = (\tau_1, \dots, \tau_n) \quad (7.6)$$

donde definimos

$$\tau_h = \xi_{i,h} - \xi_{j,h}; \quad h = 1, 2, \dots, n \quad (7.6.a)$$

Asociamos a la v.a. (7.6) el parámetro real

$$\lambda_{ij} = (\lambda_1, \dots, \lambda_n) \quad (7.6.b)$$

siendo

$$\lambda_h = \lambda_{ih} - \lambda_{jh}; \quad h = 1, 2, \dots, n \quad (7.6.c)$$

la h -ésima componente de (7.6.b), que es a su vez el parámetro real que asociamos a la v.a. (7.6.a)

Simbolizamos con

$$t_{i,j} = (t_1, \dots, t_n)$$

una particular determinación de la v.a. (7.6), correspondiendo en consecuencia a la v.a. (7.6.a)

$$t_h = x_{ih} - x_{jh} \quad (7.6.d)$$

La f. de d. de (7.6), en la hipótesis de v.a. continuas, es

$$\begin{aligned} g_{ij}(t_{ij}) = & \int_{-\infty}^{\infty} dx_{j1} \dots \int_{-\infty}^{\infty} f_{ij}(t_1 + x_{j1}, \dots, t_n + \\ & + x_{jn}, x_{j1}, \dots, x_{jn}) dx_{jn}. \end{aligned} \quad (7.7)$$

Trabajando directamente con distribuciones marginales k -dimensionales, $1 \leq k < n$, vamos a obtener los correspondientes parámetros marginales de la transvariación, cuyo caso límite,

para $k = n$, nos da los parámetros de la transvariación en el espacio total.

Para la definición de transvariación marginal k -dimensional, su intensidad de orden r y medida de dicha transvariación, para la nueva v.a. τ_{ji} ; $i \neq j$ ($i, j = 1, \dots, s$), dada en (7.6), como así también para la definición de los parámetros de la transvariación y obtención de sus correspondientes expresiones matemáticas, para los casos de v.a. dependientes, independientes, normalmente correlacionadas y normalmente independientes, nos remitimos al capítulo precedente.

7.2. — MEDIDA SINTETICA DE LA TRANSVARIACIÓN

Por los razones expuestas en 5.5 y 6.7 introducimos, como en dichos apartados, las v.a. reducidas n -dimensionales y las v.a. reducidas marginales k -dimensionales, $1 \leq k < n$.

Introducimos las v.a. $\eta_{i,h}$ ($i = 1, \dots, s$; $h = 1, \dots, n$) definidas por la expresión

$$\eta_{ih} = \frac{\xi_{i,h}}{\Delta_{hh}} \quad (7.8)$$

La v.a. η_{ih} nos define la v.a. reducida marginal unidimensional correspondiente a la h -ésima componente de la v.a. n -dimensional introducida en (7.1). Trabajando con las v.a. ξ_i y ξ_j , tenemos para η_{jh}

$$\eta_{jh} = \frac{\xi_{jh}}{\Delta_{hh}}$$

donde con Δ_{hh} simbolizamos la diferencia media entre las v.a. ξ_{ih} y ξ_{jh} . La v.a. reducida marginal k -dimensional correspondiente a la v.a. ξ_i tomando sus k primeras componentes, es, por definición

$$\eta_{(k)i} = \sum_{h=1}^k \eta_{i,h} = \sum_{h=1}^k \frac{\xi_{ih}}{\Delta_{hh}} \quad (7.9)$$

y la correspondiente a la v.a. ξ_j la obtenemos con solo reemplazar en la anterior i por j .

Simbolizando respectivamente con $u_{(k)i}$ y $u_{(k)j}$ las correspondientes determinaciones particulares de las v.a. $\eta_{(k)i}$ y $\eta_{(k)j}$ tenemos, para $u_{(k)i}$:

$$u_{(k)i} = \sum_{h=1}^k u_{i,h} = \sum_{h=1}^k \frac{x_{ih}}{\Delta_{hh}} \quad (7.9.a)$$

Para calcular $u_{(k)i}$ y $u_{(k)j}$ observamos la regla dada en 6.7.

Para la transvariación entre las v.a. reducidas marginales k -dimensional $\eta_{(k)i}$ y $\eta_{(k)j}$ nos remitimos al capítulo precedente, donde señalamos que su introducción lleva como finalidad hacer posible su tratamiento como si fuesen v.a. unidimensionales.

En lo que sigue del presente capítulo, trabajaremos con v.a. reducidas, salvo expresa indicación en contrario.

7.3. — EXPERIMENTO ALEATORIO OBJETO DE ESTUDIO

El e.a. objeto de nuestro estudio en este capítulo es el que consiste en la observación de dos elementos, no pertenecientes al mismo colectivo, del conjunto posible de s colectivos. Dicho e.a. será uno de los $s(s-1)$ posibles e.a. combinados

$$\Lambda_{ij} \ (i \neq j; i, j = 1, \dots, s)$$

Siendo

$$N = \sum_{h=1}^s N_h \quad (7.10)$$

el número total de elementos de los s colectivos considerados y partiendo de la hipótesis que los N elementos forman una serie completa de sucesos mutuamente excluyentes e igualmente posibles, la probabilidad de que las dos observaciones pertenezcan a los colectivos i -ésimo y j -ésimo, en dicho orden, es

$$\frac{N_i N_j}{N(N - N_i)} \quad (7.11)$$

La observación o extracción de un elemento cualquiera de los N posibles, implica la obtención de n valores numéricos que constituyen las particulares determinaciones, en esa observación

o extracción, de las correspondientes n componentes de la v.a. n -dimensional del colectivo al cual pertenece el elemento observado o extraído.

7.4. — PROBABILIDAD DE TRANSVARIACIÓN MARGINAL

La probabilidad de que el signo de la diferencia entre los dos valores reducidos marginales, resultado de las dos observaciones realizadas en el e.a. combinado, sea opuesto al signo de la diferencia de sus correspondientes parámetros λ , nos define la transvariabilidad marginal k -dimensional entre las s distribuciones. Simbolizándola con ${}_s\dot{p}_{(k)}$ tenemos

$${}_s\dot{p}_{(k)} = \sum_{i \neq j} \frac{N_i N_j \dot{p}_{ij}}{N (N - N_i)} \quad (7.12)$$

Con \dot{p}_{ij} simbolizamos aquí, por razones de mayor simplicidad, la transvariabilidad marginal k -dimensional entre las v.a. reducidas $\eta_{(k)i}$ y $\eta_{(k)j}$. Para el máximo de la transvariabilidad tenemos

$${}_s\dot{p}_{M(k)} = \sum_{i \neq j} \frac{N_i N_j \dot{p}_{ij;M}}{N (N - N_i)} \quad (7.13)$$

Luego

$${}_sP_{(k)} = \frac{{}_s\dot{p}_{(k)}}{{}_s\dot{p}_{M(k)}} \quad (7.14)$$

nos define la probabilidad de transvariación marginal k -dimensional para el conjunto dado de s colectivos.

Si es $N_1 = N_2 = \dots = N_s$, la (7.12), (7.13) y (7.14) resultan, respectivamente

$${}_s\dot{p}_{(k)} = \frac{2}{s(s-1)} \sum_{i < j} \dot{p}_{ij} \quad (7.15)$$

$${}_s\dot{p}_{M(k)} = \frac{2}{s(s-1)} \sum_{i < j} \dot{p}_{ij;M} \quad (7.16)$$

$${}_s\dot{p}_{(k)} = \frac{\sum_{i < j} \dot{p}_{ij}}{\sum_{i < j} \dot{p}_{ij;M}} \quad (7.17)$$

La (7.15), (7.16) y (7.17), nos dan también, respectivamente, la transvariabilidad, su valor máximo y la probabilidad de transvariación marginal k -dimensional, en el supuesto que, para observar un elemento de uno de los s colectivos dados, sea necesario primeramente seleccionar el colectivo, presentando todos ellos la misma probabilidad de ser elegido.

Si los parámetros λ son iguales a los valores medianos de sus correspondientes *v.a.* reducida, es entonces, por definición de valor mediano

$$p_{ij;M} = \frac{1}{2}$$

de donde resulta, tanto para la (7.13) como para la (7.16)

$${}_s p_{M(k)} = \frac{1}{2}$$

En consecuencia tenemos, respectivamente, para la probabilidad de transvariación dada en (7.14) y (7.17)

$${}_s P_{(k)} = 2 {}_s p_{(k)} \quad (7.18)$$

$${}_s P_{(k)} = 2 {}_s p_{(k)} = \frac{4}{s(s-1)} \sum_{i < j} p_{ij} \quad (7.19)$$

Es decir que la probabilidad de transvariación marginal k -dimensional, con respecto a sus correspondientes valores medianos — sino trabajáramos con las *v.a.* reducidas, serían hiperplanos medianos — es igual a la media aritmética de las $s(s-1)$ posibles probabilidades de transvariación marginal k -dimensional entre las *v.a.* $\eta_{(k)i}$ y $\eta_{(k)j}$ ($i \neq j$) ponderadas por la probabilidad de que las dos observaciones pertenezcan a los colectivos que ellas describen.

Es de destacar que el cálculo de $\frac{s(s-1)}{2}$ probabilidades de transvariación a que dan lugar el empleo de la (7.18) y (7.19), puede ser reducido al cálculo de $2(s-1)$ para la (7.18), cuando es $s > 4$ y de $s-1$ para la (7.19), cuando es $s > 2$.

En efecto, para la (7.18) tenemos

$${}_sP_{(k)} = \frac{2}{N} \left(\sum_{i=1}^{s-1} \frac{N_i}{N - N_i} \sum_{j=i+1}^s N_j p_{ij} + \sum_{i=2}^s \frac{N_i}{N - N_i} \sum_{j=1}^{i-1} N_j p_{ij} \right) \quad (7.20)$$

Teniendo presente que $Q_{ij} = N_i N_j p_{ij}$ nos define la frecuencia de la transvariabilidad, es decir el número de las transvariaciones posibles entre las v.a. $\eta_{(k)i}$ y $\eta_{(k)j}$ y ordenando las s variables reducidas conforme al orden de magnitud creciente o decreciente de sus correspondientes valores medianos, tenemos

$${}_sP_{(k)} = \frac{2}{N} \left[\sum_{i=1}^{s-1} \frac{N_i}{N - N_i} \sum_{h=i+1}^s N_h p_{i(i+1,\dots,s)} + \sum_{i=2}^s \frac{N_i}{N - N_i} \sum_{h=1}^{i-1} N_h p_{i(1,2,\dots,i-1)} \right] \quad (7.21)$$

donde con $p_{i(i+1,\dots,s)}$ y $p_{i(1,2,\dots,i-1)}$ simbolizamos, respectivamente, la transvariabilidad marginal entre los elementos del colectivo i -ésimo y el conjunto de elementos de los $(s-i)$ colectivos $i+1, \dots, s$ con mediana mayor (menor) y el de los $i-1$ colectivos $1, 2, \dots, i-1$ con mediana menor (mayor).

Para la (7.19), supuesto ya ordenado sus valores medianos, tenemos

$${}_sP_{(k)} = \frac{4s}{(s-1)N^2} \sum_{i < j} Q_{ij} = \frac{4}{s(s-1)} \sum_{i=1}^{s-1} (s-i) p_{i(i+1,\dots,s)} \quad (7.22)$$

Observamos también en este resultado donde hemos supuesto que los s colectivos constan del mismo número de elementos y que están ordenados conforme al orden creciente de magnitud de las medianas de sus correspondientes variables reducidas marginales k -dimensionales, que es igual a la media aritmética de la probabilidad de transvariación entre los elementos de cada colectivo con el conjunto de los elementos de los restantes colectivos con mediana mayor, ponderada por el número de estos colectivos con mediana mayor.

Una análoga interpretación de media aritmética recibe la (7.21), cuyo factor de ponderación es igual a la probabilidad de que las dos observaciones consideradas pertenezca una al colectivo i -ésimo y la otra a uno de los $s-i$ colectivos siguientes o $i-1$ colectivos precedentes, según sea al término considerado.

7.5. — INTENSIDAD DE TRANSVARIACIÓN MARGINAL

El momento de transvariación marginal k -dimensional de orden r entre las s distribuciones consideradas, es por definición igual a la esperanza matemática de las intensidades de orden r de todas las transvariaciones posibles entre las medidas de los atributos objeto de estudio de los colectivos seleccionados en el *e.a.* definido en 7.3, con respecto a sus correspondientes parámetros λ . Simbolizando este concepto con ${}_s m_{r(k)}$; con ${}_s m_{r,M(k)}$ su valor máximo y con ${}_s I_{r(k)}$ la intensidad de transvariación marginal k -dimensional de orden r , siempre para el conjunto de los s colectivos considerados, tenemos, respectivamente

$${}_s m_{r(k)} = \sum_{i \neq j} \frac{N_i N_j m_{ij;r}}{N(N - N_i)} \quad (7.23)$$

$${}_s m_{r,M(k)} = \sum_{i \neq j} \frac{N_i N_j m_{ij;r,M}}{N(N - N_i)} \quad (7.24)$$

$${}_s I_{r(k)} = \frac{({}_s m_{r(k)})^{\frac{1}{r}}}{({}_s m_{r,M(k)})^{\frac{1}{r}}} \quad (7.25)$$

Con $m_{ij;r}$ y $m_{ij;r,M}$ simbolizamos el momento de transvariación marginal y su valor máximo entre las v.a. $\eta_{(k)i}$ y $\eta_{(k)j}$.

Si los s colectivos considerados constan del mismo número de elementos la (7.25) resulta

$${}_s I_{r(k)} = \frac{(\sum_{i < j} m_{ij;r})^{\frac{1}{r}}}{(\sum_{i < j} m_{ij;r,M})^{\frac{1}{r}}} \quad (7.26)$$

que nos define también la intensidad de transvariación marginal k -dimensional de orden r , para el supuesto que su cálculo se efectuó luego de seleccionar los dos colectivos, siendo todos ellos igualmente probables, tengan o no el mismo número de elementos.

Siendo el máximo del momento de transvariación marginal de orden uno entre las v.a. reducidas $\eta_{l(k)i}$ y $\eta_{l(k)j}$, con respecto a sus medias aritméticas

$$m_{ij,1,M} = \frac{1}{2} \Delta_{ij}$$

donde con Δ_{ij} simbolizamos la diferencia media entre las v.a. reducidas consideradas, tomando como origen la media aritmética, tenemos para la intensidad de transvariación marginal k -dimensional de orden uno

$${}_s I_{1(k)} = \frac{\sum_{i \neq j} \frac{N_i N_j m_{ij,1}}{N - N_i}}{\sum_{i \neq j} \frac{N_i N_j m_{ij,1,M}}{N - N_i}} = \frac{\sum_{i \neq j} \frac{N_i N_j \Delta_{ij} I_{ij,1}}{N - N_i}}{\sum_{i \neq j} \frac{N_i N_j \Delta_{ij}}{N - N_i}} \quad (7.27)$$

Si los s colectivos son igualmente numerosos

$${}_s I_{1(k)} = \frac{\sum_{i < j} m_{ij,1}}{\sum_{i < j} m_{ij,1,M}} = \frac{\sum_{i < j} \Delta_{ij} I_{ij,1}}{\sum_{i < j} \Delta_{ij}} \quad (7.28)$$

La (7.27) y la (7.28) son medias aritméticas ponderadas de las intensidades de transvariación marginal k -dimensional de orden uno, entre las v.a. $\eta_{l(k)i}$ y $\eta_{l(k)j(i \neq j)}$, con respecto a sus correspondientes parámetros λ iguales a la media de cada una de sus componentes. El factor de ponderación de la (7.27) es la diferencia media de las dos v.a. desvíos consideradas, multiplicada por la probabilidad de que las dos observaciones seleccionadas pertenezcan a los colectivos descriptos por dichas v.a. El factor de ponderación de la (7.28) está dado por la diferencia media.

Ordenando las s variables reducidas marginales k -dimensionales conforme al orden de magnitud creciente o decreciente de sus respectivas medias aritméticas y recordando

$$I_{ij,1} = \frac{2 m_{ij,1}}{\Delta_{ij}} = \frac{2 T_{ij}}{N_i N_j \Delta_{ij}} = \frac{2 T_{ij}}{D_{ij}} \quad (7.29)$$

donde con T_{ij} simbolizamos la suma de las intensidades de transvariación de orden uno entre las v.a. $\eta_{l(k)i}$ y $\eta_{l(k)j}$ y con D_{ij} la

suma de los valores absolutos de todas sus diferencias posibles, tomando como origen la media aritmética, tenemos para la (7.27)

$$\begin{aligned}
 {}_s I_{1(k)} &= \frac{2 \left[\sum_{i=1}^{s-1} \frac{1}{N - N_i} \sum_{j=i+1}^s T_{ij} + \sum_{i=2}^s \frac{1}{N - N_i} \sum_{j=1}^{i-1} T_{ij} \right]}{\sum_{i=1}^{s-1} \frac{1}{N - N_i} \sum_{j=i+1}^s D_{ij} + \sum_{i=2}^s \frac{1}{N - N_i} \sum_{j=1}^{i-1} D_{ij}} = \\
 &= \frac{\sum_{i=1}^{s-1} \frac{N_i \sum_{j=i+1}^s N_j}{N - N_i} m_{i(i+1, \dots, s);1} + \sum_{i=2}^s \frac{N_i \sum_{j=1}^{i-1} N_j}{N - N_i} m_{i(1, \dots, i-1);1}}{\sum_{i=1}^{s-1} \frac{N_i \sum_{j=i+1}^s N_j}{N - N_i} m_{i(i+1, \dots, s);1;M} + \sum_{i=2}^s \frac{N_i \sum_{j=1}^{i-1} N_j}{N - N_i} m_{i(1, \dots, i-1);1;M}} = \quad (7.30) \\
 &= \frac{\sum_{i=1}^{s-1} \frac{N_i \sum_{j=i+1}^s N_j}{N - N_i} \Delta_{i(i+1, \dots, s)} I_{i(i+1, \dots, s);1} + \sum_{i=2}^s \frac{N_i \sum_{j=1}^{i-1} N_j}{N - N_i} \Delta_{i(1, \dots, i-1)} I_{i(1, \dots, i-1);1}}{\sum_{i=1}^{s-1} \frac{N_i \sum_{j=i+1}^s N_j}{N - N_i} \Delta_{i(i+1, \dots, s)} + \sum_{i=2}^s \frac{N_i \sum_{j=1}^{i-1} N_j}{N - N_i} \Delta_{i(1, \dots, i-1)}}
 \end{aligned}$$

que continúa siendo una media aritmética ponderada, o si se prefiere la media agregativa de dos medias aritméticas.

Con $\Delta_{i(i+1, \dots, s)}$ y con $I_{i(i+1, \dots, s);1}$, simbolizamos, respectivamente, la diferencia media y la intensidad de transvariación de orden uno entre los elementos del colectivo i -ésimo y los elementos de los $(s-i)$ colectivos con media mayor (menor) $i+1, \dots, s$. Con $m_{i(i+1, \dots, s);1}$, simbolizamos su correspondiente momento de transvariación de orden uno y con $m_{i(i+1, \dots, s);1;M}$, su valor máximo.

Para la (7.28), tenemos

$$\begin{aligned}
 {}_s I_{1(k)} &= \frac{2 \sum_{i < j} T_{ij}}{\sum_{i < j} D_{ij}} = \frac{\sum_{i=1}^{s-1} (s-i) m_{i(i+1, \dots, s);1}}{\sum_{i=1}^{s-1} (s-i) m_{i(i+1, \dots, s);1;M}} = \\
 &= \frac{\sum_{i=1}^{s-1} (s-i) \Delta_{i(i+1, \dots, s)} I_{i(i+1, \dots, s);1}}{\sum_{i=1}^{s-1} (s-i) \Delta_{i(i+1, \dots, s)}} \quad (7.31)
 \end{aligned}$$

que es la media aritmética de las intensidades de transvariación que indican sus subíndices, ponderadas por el producto entre sus correspondientes diferencias medias y el número de colectivos que agrupan.

Con la (7.30) y la (7.31), hemos reducido el cálculo de $\frac{s(s-1)}{2}$ intensidades de transvariación, al cálculo de $2(s-1)$, para la primera, cuando es $s > 4$ y de $s-1$ para la segunda, cuando es $s > 2$.

7.6. — ESPACIO DE TRANSVARIACION MARGINAL

Extendemos al conjunto de s v.a., el concepto de área y espacio de transvariación, introducido por Gini y Livada (ref. 15) para dos v.a. independientes.

Si las v.a. reducidas $\eta_{(k)i}$ y $\eta_{(k)j}$, definidas en (7.9) para $i \neq j$ son independientes, el espacio de transvariación marginal k dimensional para el e.a. definido en 7.3, es

$${}_sC_{(k)} = \sum_{i \neq j} \frac{N_i N_j}{N(N - N_i)} C_{ij} \quad (7.32)$$

donde con C_{ij} simbolizamos el espacio de transvariación marginal k -dimensional entre las v.a. $\eta_{(k)i}$ y $\eta_{(k)j}$, que es igual a la probabilidad de que en la realización del e.a. Λ_{ij} , se observe un valor $\mu_{(k)}$ perteneciente a la v.a. reducida $\eta_{(k)i}$, si es $N_i f_i(u)$ no mayor de $N_j f_j(u)$ o perteneciente a la v.a. reducida $\eta_{(k)j}$; si es $N_i f_i(u)$ mayor de $N_j f_j(u)$.

Simbolizamos con $f_i(u)$ la f. de d. de la v.a. $\eta_{(k)i}$ y con $f_j(u)$ la f. de d. de la v.a. $\eta_{(k)j}$.

Definimos el máximo del espacio de transvariación marginal k -dimensional dado en (7.32) al valor que adquiere dicha expresión en la hipótesis que las s v.a. consideradas sean idénticas y pertenezcan a colectivos con el mismo número de elementos. En base a tal hipótesis es $C_{ij} = \frac{1}{2}$. Llevando este valor a (7.32), tenemos para el máximo del espacio de transvariación

$${}_sC_{M(k)} = \frac{1}{2} \quad (7.33)$$

El cociente entre la (7.32) y la (7.33) nos define el espacio relativo de transvariación marginal k dimensional. Simbolizando este concepto con ${}_sH_{(k)}$ tenemos

$${}_sH_{(k)} = 2 {}_sC_{(k)} = \sum_{i \neq j} \frac{N_i N_j}{N(N - N_i)} H_{ij} \quad (7.34)$$

que es una media aritmética de los espacios relativos de transvariación marginal k -dimensional entre las v.a. reducidas $\eta_{(k)i}$ y $\eta_{(k)j}$, y ponderada por la probabilidad dada en (7.11)

Para $N_1 = N_2 = \dots = N_s$, resulta

$${}_sH_{(k)} = \frac{2}{s(s-1)} \sum_{i < j} H_{ij} \quad (7.35)$$

7.7. — ERROR DEL VALOR DISCRIMINATIVO MARGINAL

Sentando nuevamente la hipótesis de independendencia entre las v.a. reducidas $\eta_{(k)i}$ y $\eta_{(k)j(i \neq j)}$, tenemos para el error del valor discriminativo marginal k -dimensional para el e.a. definido en 7.3

$${}_s\varepsilon_{(k)} = \sum_{i \neq j} \frac{N_i N_j}{N(N - N_i)} \varepsilon_{ij}(z) \quad (7.36)$$

donde con $\varepsilon_{ij}(z)$ simbolizamos el error del valor discriminativo marginal k -dimensional entre las v.a. $\eta_{(k)i}$ y $\eta_{(k)j}$, siendo z el valor discriminativo marginal k -dimensional, es decir el valor z de la v.a. reducida que hace mínima la función del error.

$$\varepsilon_{ij}(x) = \frac{n_i(x) + n_j(x)}{N_i + N_j} \quad (7.37)$$

tomando la más favorable en el sentido de mínimo de las hipótesis (I) y (II) dadas respectivamente en (4.23) y (4.24).

En la hipótesis (I), $\eta_i(x)$ nos da el error que se comete al suponer que todos los valores de la v.a. reducida $\eta_{(k)i}$ son mayores de x ; $n_j(x)$ nos da el error que se comete al suponer que todos los valores de la v.a. reducida $\eta_{(k)j}$ son menores de x . Una interpretación recíproca recibe la hipótesis (II).

Si los s colectivos contienen el mismo número de elementos, la (7.36) resulta

$$s\varepsilon_{(k)} = \frac{2}{s(s-1)} \sum_{i < j} \varepsilon_{ij}(z) \quad (7.39)$$

que nos define el error del valor crítico.

7.8. — HIPOTESIS DE DISTRIBUCIONES NORMALES

7.8.1. — Si las s v.a. n -dimensionales $\xi_i = (\xi_{i,1}, \dots, \xi_{i,n})$ $i = 1 \dots s$ asociadas al e.a. Λ_i definido en 7.1, se distribuyen normalmente, resulta por lo ya visto en 5.6, que las v.a. reducidas definidas en (7.9), para $i = 1, \dots, s$ y $k = 1, \dots, n$, también se distribuyen normalmente, con media

$$M_{(k)i} = E(\eta_{(k)i}) = \sum_{h=1}^k \frac{M_{i,h}}{\Delta_{hh}} = \sum_{h=1}^k M_{(1)ih} \quad (7.40)$$

y varianza

$$\sigma_{(k)i}^2 = D^2(\eta_{(k)i}) = \sum_{h=1}^k \frac{\sigma_{ih}^2}{\Delta_{hh}^2} + 2 \sum_{h < l} \frac{\mu_{11}(\xi_{ih}, \xi_{il})}{\Delta_{hh} \Delta_{ll}} \quad (7.41.a)$$

si las n componentes de la v.a. definida en (7.1) se encuentran normalmente correlacionadas, y con

$$\sigma_{(k)i}^2 = D^2(\eta_{(k)i}) = \sum_{h=1}^k \frac{\sigma_{ih}^2}{\Delta_{hh}^2} \quad (7.41.b)$$

si son independientes.

Los sumandos de (7.40) como así también los dobles productos de (7.41.a) se tomarán con el signo que les corresponda conforme a la regla dada al introducir las v.a. reducidas en 6.7.

A los efectos del cálculo de la probabilidad e intensidad de transvariación, introducimos la nueva v.a.

$$\eta_{(k)ij} = \eta_{(k)i} - \eta_{(k)j} \quad (7.42)$$

la cual, por lo demostrado en 3.9, se distribuye normalmente con media

$$M_{(k)ij} = M_{(k)i} - M_{(k)j} \quad (7.43)$$

y varianza

$$\sigma_{(k)ij}^2 = \sigma_{(k)i}^2 + \sigma_{(k)j}^2 - 2\mu_{11}(\eta_{(k)i}, \eta_{(k)j}) \quad (7.44.a)$$

si las v.a. $\eta_{(k)i}$ y $\eta_{(k)j}$ ($i \neq j$; $i, j = 1, \dots, s$) están normalmente correlacionadas, con

$$\sigma_{(k)ij}^2 = \sigma_{(k)i}^2 + \sigma_{(k)j}^2 \quad (7.44.b)$$

si son independientes

Para d_1 renemos

$$d_1 = \frac{M_{(k)ij}}{\sigma_{(k)ij}} \quad (7.45)$$

7.8.2. — PROBABILIDAD DE TRANSVARIACION MARGINAL

La probabilidad de transvariación marginal k-dimensional entre las s v.a. reducidas en la hipótesis de distribuciones normales, trabajando con respecto a sus correspondientes valores medios, la obtenemos llevando a (7.18) o (7.19) según se trate de colectivos de diferentes o de igual tamaño, el valor de

$$p_{ij} = \Phi(-|d_1|) = \Phi\left(-\left|\frac{M_{(k)ij}}{\sigma_{(k)ij}}\right|\right) \quad (7.46)$$

deducido en 3.9. Para el cálculo abreviado de la probabilidad de transvariación, se obtiene en forma inmediata, para la (7.21) o (7.22), según corresponda, el valor de

$$p_{i(i+1, \dots, s)} = \Phi\left(-\left|\frac{M_{(k)i(i+1, \dots, s)}}{\sigma_{(k)i(i+1, \dots, s)}}\right|\right) \quad (7.46.a)$$

donde $M_{(k)i(i+1, \dots, s)}$ y $\sigma_{(k)i(i+1, \dots, s)}^2$ nos definen, respectivamente, la media y la varianza de la v.a. reducida

$$\eta_{(k)i(i+1, \dots, s)} = \eta_{(k)(i+1, \dots, s)} - \eta_{(k)i} \quad (7.47)$$

A la v.a. $\eta_{(k)(i+1, \dots, s)}$ le corresponde la f. de p.

$$f_{i(i+1, \dots, s)}(\mathcal{U}) = \frac{1}{\sum_{t=i+1}^s N_t} \sum_{t=i+1}^s N_t f_t(\mathcal{U}_{(k)t}) \quad (7.47.a)$$

Para el supuesto de que los s colectivos fuesen del mismo tamaño, se deduce para su f. de p. la expresión

$$f_{(i+1, \dots, s)}(u) = \frac{1}{s-i} \sum_{t=i+1}^s f_t(u_{(k)t}) \quad (7.47.b)$$

La v.a. reducida

$$\eta_{(k)i(1, \dots, i-1)} = \eta_{(k)(1, \dots, i-1)} - \eta_{(k)i} \quad (7.48)$$

recibe una interpretación semejante a la v.a. de (7.47). Con $M_{(k)i(1, \dots, i-1)}$ simbolizamos su media y con $\sigma_{(k)i(1, \dots, i-1)}^2$ su varianza.

7.8.3. — INTENSIDAD DE TRANSVARIACIÓN MARGINAL

Obtenemos la intensidad de transvariación marginal k -dimensional de orden r , con respecto a la media, para el e.a. dado en 7.3, en la hipótesis de distribuciones normales, reemplazando en (7.25) o (7.26), según corresponda, el momento de transvariación marginal de orden r y su valor máximo, respectivamente por

$$m_{ij,r} = (M_{(k)ij})^r \Phi(-|d_1|) + \frac{1}{\sqrt{2\pi}} \sum_{t=1}^r \binom{r}{t} (M_{(k)ij})^{r-t} \sigma_{(k)ij}^t G(t; |d_1|) \quad (7.49)$$

$$m_{ij,r,M} = \frac{2^{\frac{r-2}{2}} \sigma_{(k)ij}^r}{\sqrt{\pi}} \cdot \Gamma\left(\frac{r+1}{2}\right) \quad (7.50)$$

cuya deducción fué realizada en 3.9. El valor de d_1 , viene dado por (7.45), el de $M_{(k)ij}$ por (7.43) y el de $\sigma_{(k)ij}$ por la raíz cuadrada de la (7.44.a) o de la (7.44.b) según que las v.a. reducidas se encuentren normalmente correlacionadas o sean independientes. El valor de $G(t; |d_1|)$ para t par e impar, lo obtenemos a partir de (2.48) y (2.47).

Si llevamos a (7.27) y (7.28) los valores de (7.49) y (7.50) luego de hacer $r = 1$, tendremos la intensidad de transvariación marginal k -dimensional de orden uno, para el supuesto, respectivamente, de que los s colectivos normales que intervienen en el e.a. dado en 7.3, sean de diferente o de igual tamaño.

Reemplazando en (7.49) y (7.50), luego de hacer $r = \bar{r}$, la media $M_{(k)ij}$ por $M_{(k)i(i+1, \dots, s)}$ o por $M_{(k)i(1, \dots, i-1)}$ el desvío medio cuadrático $\sigma_{(k)ij}$ por $\sigma_{(k)i(i+1, \dots, s)}$ o por $\sigma_{(k)i(1, \dots, i-1)}$ de donde resulta, para cada caso

$$d_1 = \frac{M_{(k)i(i+1, \dots, s)}}{\sigma_{(k)i(i+1, \dots, s)}} \quad (7.51.a)$$

y

$$d_1 = \frac{M_{(k)i(1, \dots, i-1)}}{\sigma_{(k)i(1, \dots, i-1)}} \quad (7.51.b)$$

tendremos los valores de $m_{i(i+1, \dots, s);1}$ y $m_{i(i+1, \dots, s);1;M}$ y los de $m_{i(1, \dots, i-1);1}$ y $m_{i(1, \dots, i-1);1;M}$, que llevados a (7.30) y (7.31), nos darán la intensidad de transvariación marginal k -dimensional de orden uno, por el método abreviado de trabajo, para el sistema de hipótesis consideradas en este parágrafo.

7.8.4. — ESPACIO DE TRANSVARIACION MARGINAL

Si las v.a. reducidas definidas en (7.9) para $i = 1, 2, \dots, s$, son independientes dos a dos, obtenemos el espacio relativo de transvariación marginal k -dimensional, para el e.a. definido en 7.3, reemplazando H_{ij} en (7.34) o (7.35), según corresponda, por la expresión

$$H_{ij} = \frac{2}{N_i + N_j} \int_{-\infty}^{\infty} \delta_{ij}(u) du \quad (7.52)$$

siendo $\delta_{ij}(u)$ igual a una de las siguientes expresiones

$$\delta_{ij}(u) = \begin{cases} N_i f_i(u; M_{(k)i}, \sigma_{(k)i}) = \frac{N_i}{\sigma_{(k)i} \sqrt{2\pi}} e^{-\frac{(u-M_{(k)i})^2}{2\sigma_{(k)i}^2}} \\ N_j f_j(u; M_{(k)j}, \sigma_{(k)j}) = \frac{N_j}{\sigma_{(k)j} \sqrt{2\pi}} e^{-\frac{(u-M_{(k)j})^2}{2\sigma_{(k)j}^2}} \end{cases} \quad (7.53)$$

según sea

$$N_i f_i(u; M_{(k)i}; \sigma_{(k)i}) \leq N_j f_j(u; M_{(k)j}; \sigma_{(k)j}) \quad (7.54.a)$$

o bien

$$N_i f_i(u; M_{(k)i}; \sigma_{(k)i}) > N_j f_j(u; M_{(k)j}; \sigma_{(k)j}) \quad (7.54.b)$$

7.8.5. ERROR DEL VALOR DICRIMINATIVO MARGINAL

Para determinar el error del valor discriminativo marginal k -dimensional, para el mismo sistema de hipótesis dado en 7.8.4, reemplazamos primeramente en (4.23) y (4.24) la f. de p. $f_1(t)$ por la correspondiente a la v.a. reducida normal $M_{(k)i}; \sigma_{(k)i}$ y $f_2(t)$ por la f. de p. de la v.a. reducida normal $M_{(k)j}; \sigma_{(k)j}$ con lo que obtenemos las funciones $n_i(x)$ y $n_j(x)$. Llevando estas expresiones a (7.37) deducimos el valor z de x para el cual presenta un mínimo la función $\xi_{ij}(x)$ tomando la más favorable, en el sentido de mínimo de las hipótesis contenidas en (4.23) y (4.24). Reemplazando en (7.36) o en (7.39), según corresponda, los valores $\xi_{ij}(z)$ para todas las combinaciones binarias de i con j , tenemos el error del valor discriminativo marginal k -dimensional para el e.a. definido en 7.3.

8. — VARIANZA Y COVARIANZA DE LOS PARAMETROS DE LA TRANSVARIACION EN EL MUESTREO

8.1. — INTRODUCCION

En los capítulos precedentes nos hemos ocupado de los parámetros de la transvariación, deduciéndolos para el sistema de hipótesis considerado en cada uno de ellos. En el campo de las aplicaciones, estos parámetros deben ser estimados a través del conjunto de observaciones obtenidas mediante el muestreo estadístico, razón por la cual dichos valores están sometidos a las fluctuaciones de la muestras, definiéndonos en ella una nueva v.a., de donde se desprende la importancia de conocer para grandes muestras, la media y la varianza de tales v.a., para la estimación de cada parámetro de la transvariación.

Para las grandes muestras podemos hacer jugar con éxito, en las condiciones exigidas para cada caso, los importantes teoremas de la convergencia en probabilidad, en particular el teorema de Tchebycheff que nos dice en qué condiciones una función de una v.a., que es otra v.a., converge en probabilidad a una constante y el teorema central del límite, que nos dice en qué condiciones la distribución de una función de una v.a. converge en probabilidad a la distribución normal.

Pero la limitación en la aplicación de las capitales conclusiones de los teoremas de convergencia es evidente, desde que por muchas razones el investigador está con frecuencia imposibilitado de disponer de muestras suficientemente grandes. A suplir esta limitación concurren los modelos teóricos de distribución de las características muestrales, llamadas distribuciones exactas en el muestreo, de fecunda aplicación en las pequeñas muestras, como la t de Student, χ^2 de Pearson, Z de Fisher y F de Snedecor, la segunda Z de Fisher sobre la distribución del coeficiente de correlación de las muestras provenientes de colectivos normales, etc.

La varianza y covarianza de la probabilidad de transvariación y de la intensidad de transvariación de orden r , correspondiente a cada clase de e.a. considerado en los capítulos precedentes, serán deducidas conforme al siguiente sistema de hipótesis :

a) las observaciones que componen las muestras son obtenidas según el esquema de Bernouilli ;

b) los parámetros λ son datos conocidos a priori ; por lo tanto son independientes de las fluctuaciones en el muestreo ; en consecuencia, si trabajamos con la media o la mediana, se entenderá que son la media o la mediana del colectivo ;

c) las distribuciones de las observaciones en el muestreo se han reducido convenientemente a intervalos igualmente espaciados, con la misma amplitud en los subintervalos e igual recorrido para todas las muestras ;

d) en el capítulo 7, donde estudiamos la transvariación entre varias distribuciones, la probabilidad e intensidad de transvariación entre las distintas combinaciones binarias de las distribuciones sometidas a estudio son entre si independientes. En consecuencia, es nula la covarianza entre las distintas combinaciones binarias de la transvariabilidad, de su máximo y de es-

tos entre sí, cuando corresponden a distintas muestras; análogo resultado corresponde a las componentes de la intensidad de transvariación.

8.2. — LOS PARAMETROS DE LA TRANSVARIACIÓN EN EL CAMPO DISCRETO

La v.a. bidimensional introducida en 3.1

$$\xi = (\xi_1, \xi_2) \quad (8.1)$$

asociada al e.a. Λ , de la cual es

$$z_e = (x_i, y_j); \quad i = 1, 2, \dots, k_1; \quad j = 1, 2, \dots, k_2 \quad (8.1.a)$$

una particular determinación de la (8.1), en la hipótesis de distribución discreta, le corresponde la f . de d . $p_{i,j}$.

Siendo

$$\begin{cases} x_{i+1} - x_i = h_1; \quad i = 1, 2, \dots, k_1 \\ y_{j+1} - y_j = h_2; \quad j = 1, 2, \dots, k_2 \end{cases} \quad (8.2)$$

podemos, mediante una adecuada sustitución de variables, transformarlas en v.a. con igual subintervalo h . Introduciendo donde corresponda, subintervalos de las v.a. ξ_1 y ξ_2 con probabilidad nula, podemos hacerlas variar en los mismos intervalos. Es decir que tenemos, luego de realizar estas transformaciones y agregados convenientes, manteniendo las mismas letras:

$$x_{i+1} - x_i = y_{j+1} - y_j = h; \quad i, j = 1, 2, \dots, k \quad (8.3)$$

Suponiendo, sin pérdida de generalidad, que $\lambda_1 > \lambda_2$ y recordando la convención adoptada de considerar como transvariación la mitad de las diferencias nulas $y_j - x_i$, tenemos para la transvariabilidad, máximo de la transvariabilidad, momento de transvariación de orden r y máximo de dicho momento, respectivamente

$$p = P(\xi_1 < \xi_2) + \frac{1}{2} P(\xi_1 = \xi_2) = \sum_{j=1}^k \sum_{i=1}^{l-1} p_{i,j} + \frac{1}{2} \sum_{i=1}^k p_{ii} \quad (8.4)$$

$$\begin{aligned}
 p_M &= P(\xi_1 < \xi_2 + hc) + \frac{1}{2} P(\xi_1 = \xi_2 + hc) = \\
 &= \sum_{j=l}^k \sum_{i=1}^{l-1} p_{i,j-c} + \frac{1}{2} \sum_{i=1}^k p_{i,i-c}
 \end{aligned} \tag{8.5}$$

$$m_r = \sum_{j=l}^k \sum_{i=1}^{l-1} (y_j - x_i)^r p_{ij} = h^r \sum_{j=l}^k \sum_{i=1}^{l-1} \left(\frac{y_j - x_i}{h} \right)^r p_{ij} \tag{8.6}$$

$$m_{r,M} = \sum_{j=l}^k \sum_{i=1}^{l-1} (y_j - x_i + hc)^r p_{i,j-c} = h^r \sum_{j=l}^k \sum_{i=1}^{l-1} \left(\frac{y_j - x_i}{h} + c \right)^r p_{i,j-c} \tag{8.7}$$

Donde con c simbolizamos el entero más próximo por defecto o por exceso a $\frac{\lambda_1 - \lambda_2}{h}$. Luego

$$c = \frac{\lambda_1 - \lambda_2}{h} + \varepsilon \quad ; \quad |\varepsilon| \leq 0,50 \tag{8.8}$$

A partir de (8.4) y (8.5), obtenemos para la probabilidad de transvariación

$$P = \frac{p}{p_M} = \frac{\sum_{j=l}^k \sum_{i=1}^{l-1} p_{ij} + \frac{1}{2} \sum_{i=1}^k p_{ii}}{\sum_{j=l}^k \sum_{i=1}^{l-1} p_{i,j-c} + \frac{1}{2} \sum_{i=1}^k p_{i,i-c}} \tag{8.9}$$

A partir de (8.6) y (8.7), obtenemos para la intensidad de transvariación de orden r

$$I_r = \left(\frac{m_r}{m_{r,M}} \right)^{\frac{1}{r}} = \frac{\left[\sum_{j=l}^k \sum_{i=1}^{l-1} \left(\frac{y_j - x_i}{h} \right)^r p_{ij} \right]^{\frac{1}{r}}}{\left[\sum_{j=l}^k \sum_{i=1}^{l-1} \left(\frac{y_j - x_i}{h} + c \right)^r p_{i,j-c} \right]^{\frac{1}{r}}} \tag{8.10}$$

A los efectos de simplificar el trabajo operatorio, introducimos la transformación lineal dada en (3.16)

$$\eta = \xi_2 - \xi_1 \tag{8.11}$$

siendo

$$t_s = y_j - x_i; i, j = 1, 2, \dots, k \quad (8.11.a)$$

una particular determinación de la v.a. η .

En consecuencia, la v.a. de (8.11) puede asumir cualquier valor de la progresión aritmética

$$-(k-1)h, -(k-2)h, \dots, -h, 0, h, \dots, (k-2)h, (k-1)h$$

por ser

$$y_j - x_i = (j - i)h; i, j = 1, 2, \dots, k$$

Luego es

$$t = \frac{t_s}{h} = \frac{y_{i+t} - y_i}{h}; t = -(k-1), \dots, -1, 0, \dots, k-1 \quad (8.12)$$

siendo

$$p_t = \sum_{i=1}^{k-t} p_{i,i+t} = \sum_{i=1}^k p_{i,i+t} \quad (8.13)$$

la probabilidad de presentación del valor t . De donde

$$\sum_{t=-(k-1)}^{k-1} p_t = 1 \quad (8.13.a)$$

8.3. — DISTRIBUCION EN EL MUESTREO DE LA V.A. α_t ASOCIADA AL E.A. Λ

Dado el e.a. $\Lambda_i; i = 1, 2, \dots, n$, consistente en la observación de un elemento del colectivo descrito por la v.a. ξ_1 y otro del colectivo descrito por la v.a. ξ_2 . Se repite n veces dicho e.a. en condiciones completamente homogéneas, obteniendo así una muestra de n pares de observaciones, siguiendo el esquema de Bernouilli. Simbolizamos con $\Lambda = (\Lambda_1, \dots, \Lambda_n)$ este e.a. combinado y con

$$a_t = \sum_{i=1}^{k-t} a_{i,i+t} \quad (8.14)$$

la frecuencia correspondiente al suceso E_t ; $t = -(k-1), \dots, -1, 0, 1, \dots, k-1$ que tiene probabilidad p_t de presentarse, en la realización de dicho e.a. combinado, siendo

$$t = \frac{y_{i+t} - x_i}{h}; i = 1, 2, \dots, k-t \quad (8.14.a)$$

$$a_t = n p_t \dots \sum_{i=-(k-1)}^{k-1} a_t = n \sum_{i=-(k-1)}^{k-1} p_t = n \quad (8.14.b)$$

En la realización reiterada del e.a. combinado Λ , la frecuencia a_t describe la v.a. α_t , que es una de las $2(k-1)$ componentes del punto o vector variable

$$\alpha = (\alpha_{-(k-1)}, \dots, \alpha_{-1}, \alpha_0, \alpha_1, \dots, \alpha_{k-2}) \quad (8.15)$$

que nos define una v.a. en el espacio de $2(k-1)$ dimensiones, la que tiene una distribución polinomial en dicho espacio, en virtud de nuestra hipótesis a) de 8.1 y de la definición del e.a. combinado Λ .

La f . de d . correspondiente a la v.a. polinomial definida en (8.15) es:

$$\begin{aligned} f(\alpha) &= f(\alpha_{-(k-1)}, \dots, \alpha_{-1}, \alpha_0, \alpha_1, \dots, \alpha_{k-2}) = \\ &= \frac{n!}{\prod_{i=-(k-1)}^{k-1} \alpha_i!} \prod_{i=-(k-1)}^{k-1} p_i^{\alpha_i} \end{aligned} \quad (8.16)$$

Siendo

$$\sum_{i=-(k-1)}^{k-1} \alpha_i = n; \quad \alpha_i = 0, 1, 2, \dots, n. \quad (8.16.a)$$

$$\sum_{i=-(k-1)}^{k-1} p_i = 1. \quad (8.16.b)$$

La $f.c.$ correspondiente a la v.a. α es

$$\begin{aligned} \Phi_\alpha(u) &= \Phi(u_{-(k-1)}, \dots, u_{-1}, u_0, u_1, \dots, u_{k-2}) = E(e^{iu\alpha}) = \\ &= E[\exp(i u_{-(k-1)} \alpha_{-(k-1)} + \dots + i u_{-1} \alpha_{-1} + i u_0 \alpha_0 + i u_1 \alpha_1 + \\ &+ \dots + i u_{k-2} \alpha_{k-2})] = (p_{-(k-1)} e^{iu_{-(k-1)}} + \dots + p_{k-1} e^{iu_{k-1}})^n \end{aligned} \quad (8.17)$$

Siendo

$$u = (u_{-(k-1)}, \dots, u_{-1}, u_0, u_1, \dots, u_{k-2}) \quad (8.17.b)$$

un vector arbitrario en el espacio de $2(k-1)$ dimensiones.

Directamente, por definición de momentos, o a través de (8.17) derivando parcialmente con respecto a u_t , resulta:

$$E(\alpha_t) = \frac{1}{i} \Phi'_\alpha(0) = n p_t = a_t; \quad (8.18)$$

$$t = -(k-1), \dots, -1, 0, 1, \dots, k-1$$

Introduciendo la v.a. desvío

$$\gamma_t = \alpha_t - a_t = \alpha_t - n p_t \quad (8.19)$$

se deduce para la v.a.

$$\gamma = (\gamma_{-(k-1)}, \dots, \gamma_{-1}, \gamma_0, \gamma_1, \dots, \gamma_{k-2}) \quad (8.19.a)$$

la f.c.

$$\begin{aligned} \Phi_\gamma(u) &= E(e^{iu\gamma}) = E \left[e^{\sum_{t=-(k-1)}^{k-2} u_t (\alpha_t - n p_t)} \right] = \\ &= \left(\sum_{t=-(k-1)}^{k-1} p_t e^{iu_t - i \sum_{t=-(k-1)}^{k-2} p_t u_t} \right)^n = e^{-in \sum_{t=-(k-1)}^{k-2} p_t u_t} \Phi_\alpha(u) \end{aligned} \quad (8.20)$$

A partir de (8.20) o directamente, por definición de momentos, obtenemos, para los distintos valores posibles de t .

$$E(\gamma_t) = \Phi'_\gamma(0) = 0 \quad (8.21.a)$$

$$E(\gamma_t^2) = \Phi''_\gamma(0) = n p_t q_t \quad (8.21.b)$$

$$E(\gamma_t \gamma_v) = -n p_t p_v \quad (8.21.c)$$

8.4. — LOS PARAMETROS DE LA TRANSVARIACION EN EL MUESTREO

Trabajando con la v.a. reducida dada en (8.11) y la frecuencia α_t , correspondiente a la (8.14), tenemos, para la transvariabilidad, máximo de la transvariabilidad, momento de trans-

variación de orden r , su valor máximo, probabilidad de transvariación e intensidad de transvariación de orden r , en la realización del e.a. combinado Λ respectivamente, las siguientes expresiones

$$p_{\Delta} = \frac{1}{n} \sum_{t=1}^{k-1} \alpha_t + \frac{1}{2n} \alpha_0 \quad (8.22)$$

$$p_{M,\Delta} = \frac{1}{n} \sum_{t=-(c-1)}^{k-1} \alpha_t + \frac{1}{2n} \alpha_{-c} \quad (8.23)$$

$$m_{r,\Delta} = \frac{h^r}{n} \sum_{t=1}^{k-1} t^r \alpha_t \quad (8.24)$$

$$m_{r,M,\Delta} = \frac{h^r}{n} \sum_{t=-(c-1)}^{k-1} (t+c)^r \cdot \alpha_t \quad (8.25)$$

$$P_{\Delta} = \frac{\sum_{t=1}^{k-1} \alpha_t + \frac{1}{2} \alpha_0}{\sum_{t=-(c-1)}^{k-1} \alpha_t + \frac{1}{2} \alpha_{-c}} \quad (8.26)$$

$$\bar{I}_{r,\Delta} = \frac{\left(\sum_{t=1}^{k-1} t^r \alpha_t \right)^{\frac{1}{r}}}{\left(\sum_{t=-(c-1)}^{k-1} (t+c)^r \alpha_t \right)^{\frac{1}{r}}} \quad (8.27)$$

que nos definen, en la realización reiterada del e.a. combinado Λ , una nueva v.a. con $E(\alpha_t) = a_t = n p_t$, deducida en (8.18) y

$$E(p_{\Delta}) = \frac{1}{n} \sum_{t=1}^{k-1} E(\alpha_t) + \frac{1}{2n} E(\alpha_0) = \sum_{t=1}^{k-1} p_t + \frac{1}{2} p_0 = p \quad (8.28)$$

$$E(p_{M,\Delta}) = p_M \quad (8.29)$$

$$E(m_{r,\Delta}) = \frac{h^r}{n} \sum_{t=1}^{k-1} t^r E(\alpha_t) = h^r \sum_{t=1}^{k-1} t^r p_t = m_r \quad (8.30)$$

$$E(m_{r,M,\Delta}) = h^r \sum_{t=-(c-1)}^{k-1} (t+c)^r p_t = m_{r,M} \quad (8.31)$$

Para obtener la esperanza matemática de (8.26) y (8.27), diferenciamos logarítmicamente, con lo que tenemos, para la probabilidad de transvariación en el muestreo dada por (8.26).

$$\frac{d P_{\Delta}}{P} = \frac{d p_{\Delta}}{p} - \frac{d p_{M,\Delta}}{p_M} \quad (8.32)$$

donde las diferenciales nos definen el desvío entre el valor obtenido en la muestra y el valor poblacional.

Tomando esperanza matemática en (8.32) y teniendo presente (8.28) y (8.29), resulta :

$$E (d P_{\Delta}) = E (P_{\Delta} - P) = 0$$

Luego

$$E (P_{\Delta}) = P \quad (8.33)$$

Procediendo en forma análoga con (8.27), tenemos

$$\frac{E (d I_{r,\Delta})}{I_r} = \frac{1}{r} \left[\frac{E (d m_{r,\Delta})}{m_r} - \frac{E (d m_{r,M,\Delta})}{m_{r,M}} \right] = 0$$

por ser, en virtud de (8.30) y (8.31)

$$E (d m_{r,\Delta}) = 0$$

$$E (d m_{r,M,\Delta}) = 0$$

resulta

$$E (d I_{r,\Delta}) = E (I_{r,\Delta}) - I_r = 0$$

En consecuencia

$$E (I_{r,\Delta}) = I_r \quad (8.34)$$

Teniendo presente los resultados obtenidos en (8.18), (8.28), (8.29), (8.30), (8.31) (8.33) y (8.34), concluimos que las v.a. correspondientes a las componentes de (8.15) y a las dadas en (8.22), (8.23), (8.24), (8.25), (8.26) y (8.27), convergen en probabilidad a sus correspondientes parámetros poblacionales, el decir que sus estimaciones muestrales son estimaciones insesgadas o no viciadas.

Simbolizando con

$$\delta_p, \delta p_M, \delta m_r, \delta m_{r,M}, \delta_P \text{ y } \delta_{I_r},$$

las v.a. desvíos correspondientes a las v.a. muestrales que estamos considerando, además de γ_t definida en (8.19), tenemos

$$\begin{aligned} \delta_p &= p_\Delta - p = \frac{1}{n} \sum_{t=1}^{k-1} (\alpha_t - n p_t) + \frac{1}{2n} (\alpha_0 - n p_0) = \\ &= \frac{1}{n} \left(\sum_{t=1}^{k-1} \gamma_t + \frac{1}{2} \gamma_0 \right) \end{aligned} \quad (8.35)$$

$$\delta p_M = p_{M,\Delta} - p_M = \frac{1}{n} \left(\sum_{t=-(c-1)}^{k-1} \gamma_t + \frac{1}{2} \gamma_{-c} \right) \quad (8.36)$$

$$\delta m_r = m_{r,\Delta} - m_r = \frac{h^r}{n} \sum_{t=1}^{k-1} t^r (\alpha_t - n p_t) = \frac{h^r}{n} \sum_{t=1}^{k-1} t^r \gamma_t \quad (8.37)$$

$$\delta m_{r,M} = m_{r,M,\Delta} - m_{r,M} = \frac{h^r}{n} \sum_{t=-(c-1)}^{k-1} (t+c)^r \gamma_t \quad (8.38)$$

$$\delta_P = P_\Delta - P \quad (8.39)$$

$$\delta_{I_r} = I_{r,\Delta} - I_r \quad (8.40)$$

8.5. — MOMENTOS MIXTOS DE LAS V.A. DE LA TRANSVARIACION

Teniendo presente (8.21.b), (8.21.c) y (8.35), resulta

$$\begin{aligned} E(\delta_p \gamma_0) &= \frac{1}{n} \left[\sum_{t=1}^{k-1} E(\gamma_0 \gamma_t) + \frac{1}{2} E(\gamma_0^2) \right] = \\ &= -p_0 \sum_{t=1}^{k-1} p_t + \frac{1}{2} p_0 q_0 = \frac{1}{2} p_0 (q - p) \end{aligned} \quad (8.41)$$

$$E(\delta_p \gamma_t) = \frac{1}{n} \left[\sum_{\substack{v=1 \\ v \neq t}}^{k-1} E(\gamma_t \gamma_v) + E(\gamma_t^2) + \frac{1}{2} E(\gamma_0 \gamma_t) \right] = p_t q \quad (8.42)$$

$$E(\delta_p \gamma_{-c}) = \frac{1}{n} \left[\sum_{t=1}^{k-1} E(\gamma_{-c} \gamma_t) + \frac{1}{2} E(\gamma_{-c}^2 \gamma_0) \right] = -p_{-c} p \quad (8.43)$$

Trabajando en forma análoga con (8.36), tenemos

$$E (\delta_{p_M} \gamma_{-c}) = \frac{1}{2} p_{-c} (q_M - p_M) \quad (8.44)$$

$$E (\delta_{p_M} \gamma_t) = p_t q_M = p_t (1 - p_M) \quad (8.45)$$

Los momentos de orden dos de las v.a. definidas en (8.35) y (8.36) resultan

$$\begin{aligned} E (\delta_p^2) &= \frac{1}{n} \left[\sum_{t=1}^{k-1} E (\gamma_t \delta_p) + \frac{1}{2} E (\gamma_0 \delta_p) \right] = \\ &= \frac{1}{n} \left[q \sum_{t=1}^{k-1} p_t + \frac{1}{4} p_0 (q - p) \right] = \frac{1}{n} \left(p q - \frac{1}{4} p_0 \right) \end{aligned} \quad (8.46)$$

$$\begin{aligned} E (\delta_{p_M}^2) &= \frac{1}{n} \left[\sum_{t=-(c-1)}^{k-1} E (\gamma_t \delta_{p_M}) + \frac{1}{2} E (\gamma_{-c} \delta_{p_M}) \right] = \\ &= \frac{1}{n} \left(p_M q_M - \frac{1}{4} p_{-c} \right) \end{aligned} \quad (8.47)$$

$$\begin{aligned} E (\delta_p \delta_{p_M}) &= \frac{1}{n} \left[\sum_{t=1}^{k-1} E (\gamma_t \delta_{p_M}) + \frac{1}{2} E (\gamma_0 \delta_{p_M}) \right] = \\ &= \frac{1}{n} \left[\sum_{t=-(c-1)}^{k-1} E (\gamma_t \delta_p) + \frac{1}{2} E (\gamma_{-c} \delta_p) \right] = \frac{1}{n} p q_M \end{aligned} \quad (8.48)$$

En efecto, a partir del segundo miembro de (8.48) y teniendo presente (8.45), resulta

$$E (\delta_p \delta_{p_M}) = \frac{1}{n} \left\{ \sum_{t=1}^{k-1} p_t q_M + \frac{1}{2} p_0 q_M \right\} = \frac{1}{n} p q_M \quad (8.48.a)$$

Si trabajamos con el tercer miembro de (8.48), tenemos

$$\begin{aligned} E (\delta_p \delta_{p_M}) &= \frac{1}{n} \left\{ \sum_{t=-(c-1)}^{k-1} \frac{1}{n} \left[\sum_{v=1}^{k-1} E (\gamma_v \gamma_t) + \frac{1}{2} E (\gamma_0 \gamma_t) \right] + \right. \\ &\quad \left. + \frac{1}{2} E (\gamma_{-c} \delta_p) \right\} = \end{aligned} \quad (8.48.b)$$

$$= \frac{1}{n} \left\{ \sum_{t=-(c-1)}^0 \frac{1}{n} \sum_{v=1}^{k-1} E(\gamma_t \gamma_v) + \sum_{t=1}^{k-1} \frac{1}{n} \left[\sum_{\substack{v=1 \\ v \neq t}}^{k-1} E(\gamma_v \gamma_t) + E(\gamma_t^2) \right] \right\} + \\ + \frac{1}{n} \left\{ \frac{1}{2} \left[E(\delta_{p_M} \gamma_0) - \frac{1}{2n} E(\gamma_0 \gamma_{-c}) \right] + \frac{1}{2} E(\gamma_{-c} \delta_p) \right\} = \frac{1}{n} p q_M$$

A partir de (8.21.c) y (8.37) y (8.38) respectivamente, obtenemos

$$E(\delta_{m_r} \gamma_t) = \frac{h^r}{n} \left[\sum_{\substack{v=1 \\ v \neq t}}^{k-1} v^r E(\gamma_v \gamma_t) + t^r E(\gamma_t^2) \right] = \quad (8.49)$$

$$= h^r \left(-p_t \sum_{v=1}^{k-1} v^r p_v + t^r p_t \right) = p_t [(h t)^r - m_r] \quad (8.50)$$

$$E(\delta_{m_{r,M}} \gamma_t) = p_t [h^r (t + c)^r - m_{r,M}]$$

Para los momentos centrados de orden dos de las v.a. momento de transvariación de orden r y su máximo, se deduce, teniendo presente (8.49) y (8.50).

$$E(\delta_{m_r}^2) = \frac{h^r}{n} \sum_{t=1}^{k-1} t^r E(\gamma_t \delta_{m_r}) = \quad (8.51)$$

$$= \frac{h^{2r}}{n} \sum_{t=1}^{k-1} t^{2r} p_t - \frac{h^r}{n} m_r \sum_{t=1}^{k-1} t^r p_t = \frac{m_{2r} - m_r^2}{n}$$

$$E(\delta_{m_{r,M}}^2) = \frac{h^r}{n} \sum_{t=-(c-1)}^{k-1} (t + c)^r E(\gamma_t \delta_{m_{r,M}}) = \frac{m_{2r,M} - m_{r,M}^2}{n} \quad (8.52)$$

$$E(\delta_{m_r} \delta_{m_{r,M}}) = \frac{h^r}{n} \sum_{t=-(c-1)}^{k-1} (t + c)^r E(\gamma_t \delta_{m_r}) = \\ = \frac{h^r}{n} \sum_{t=1}^{k-1} t^r E(\gamma_t \delta_{m_{r,M}}) = \quad (8.53)$$

$$\frac{\sum_{s=0}^r \binom{r}{s} (h c)^s m_{2r-s} - m_r m_{r,M}}{n} = \frac{m_{r+r,M} - m_r m_{r,M}}{n}$$

donde convenimos en simbolizar con

$$m_{r+r,M} = E [(t+c)^r m_r] = h^{2r} \sum_{t=1}^{h-1} t^r (t+c)^r p_t = \quad (8.53.a)$$

$$= \sum_{s=0}^r \binom{r}{s} (h c)^s m_{2r-s}$$

a los efectos de mantener la correspondencia de la notación con la que se nos presenta en forma más elemental en la covarianza de los momentos muestrales.

Para la obtención del resultado de (8.53) seguimos un procedimiento similar al expuesto para (8.48) en (8.48.a) y (8.48.b)

8.6. — VARIANZA Y COVARIANZA DE LA TRANSVARIABILIDAD Y DEL MÁXIMO DE LA TRANSVARIABILIDAD

Las fórmulas deducidas en (8.46), (8.47) y (8.48), nos dan, respectivamente, una estimación sesgada o viciada de la varianza de la transvariabilidad, de su máximo y de la covarianza entre estas dos v.a.. Si reemplazamos n , tamaño de la muestra, por $n-1$, número de grados de libertad, tendremos una estimación insesgada o no viciada. Luego, la varianza de la transvariabilidad resulta

$$D^2(p_\Lambda) = \sigma_{p_\Lambda}^2 = \frac{n}{n-1} E(\delta_p^2) = \frac{1}{n-1} \left(p q - \frac{1}{4} p_0 \right) \quad (8.54)$$

La varianza del máximo de la transvariabilidad

$$D^2(p_{M,\Lambda}) = \sigma_{p_{M,\Lambda}}^2 = \frac{1}{n-1} \left(p_M q_M - \frac{1}{4} p_{-c} \right) \quad (8.55)$$

La covarianza entre las v.a. de la transvariabilidad y de su máximo

$$\mu_{11}(p_\Lambda, p_{M,\Lambda}) = \frac{n}{n-1} E(\delta_p \delta_{p_m}) = \frac{1}{n-1} p q_M \quad (8.56)$$

En efecto, tomando uno cualquiera de estos tres resultados, pues el razonamiento es semejante para ellos, tenemos, a partir

de (8.56), teniendo presente que la covarianza nos define en el muestreo la nueva v.a.

$$\mu'_{11} (p_{\Lambda}, p_{M,\Lambda}) = \frac{1}{n-1} p_{\Lambda} q_{M,\Lambda} = \frac{1}{n-1} (p_{\Lambda} - p_{\Lambda} p_{M,\Lambda})$$

y recordando los resultados obtenidos en (8.28) y (8.48)

$$\begin{aligned} E (\mu'_{11}) &= \frac{1}{n-1} [E (p_{\Lambda}) - E (p_{\Lambda} p_{M,\Lambda})] = \\ &= \frac{1}{n-1} [p - E (\delta_p \delta_{p_M}) - p p_M] = \frac{1}{n} p q_M \end{aligned}$$

8.7. — VARIANZA Y COVARIANZA DEL MOMENTO DE TRANSVARIACIÓN DE ORDEN r Y DEL MÁXIMO DE DICHO MOMENTO.

Análogamente se demuestra también que los resultados obtenidos en (8.51), (8.52) y (8.53), constituyen una estimación sesgada de la varianza del momento de transvariación de orden r , del máximo del momento de transvariación de orden r y de la covarianza entre estas dos v.a. en el muestreo.

Reemplazando el tamaño n de la muestra por el número $n - 1$ de grados de libertad, obtenemos una estimación insesgada, resultando, para la varianza del momento de transvariación de orden r

$$D^2 (m_{r,\Lambda}) = \sigma_{m_{r,\Lambda}}^2 = \frac{n}{n-1} E (\delta_{m_r}^2) = \frac{m_{2r} - m_r^2}{n-1} \quad (8.57)$$

Para la varianza del máximo del momento de transvariación de orden r

$$D^2 (m_{r,M,\Lambda}) = \sigma_{m_{r,M,\Lambda}}^2 = \frac{m_{2r,M} - m_{r,M}^2}{n-1} \quad (8.58)$$

Para la covarianza

$$\mu_{11} (m_{r,\Lambda}, m_{r,M,\Lambda}) = \frac{m_{r+r,M} - m_r m_{r,M}}{n-1} \quad (8.59)$$

donde $m_{r+r,M}$ está dado por (8.53.a)

Tomando la varianza del momento de transvariación de orden r , veremos a continuación que ella es una estimación insesgada. En forma análoga se procede con (8.58) y (8.59).

En efecto, simbolizando con $\sigma_{m_{r,\Lambda}}'^2$ la v.a. en el muestreo de la varianza del momento de transvariación de orden r , es

$$\sigma_{m_{r,\Lambda}}'^2 = \frac{m_{2r,\Lambda} - m_{r,\Lambda}^2}{n - 1}$$

y teniendo presente (8.30), (8.37) y (8.51), se deduce

$$E(\sigma_{m_{r,\Lambda}}'^2) = \frac{1}{n - 1} [E(m_{2r,\Lambda}) - E(\delta_{m_r}^2) - m_r^2] = \frac{m_{2r} - m_r^2}{n}$$

8.8. — VARIANZA DE LA PROBABILIDAD DE TRANSVARIACION.

Elevando al cuadrado ambos miembros de (8.32), que es la diferencial logarítmica de la (8.26), que nos define en el muestreo la v.a. probabilidad de transvariación para la v.a. definida en (8.11), tenemos, luego de tomar esperanza matemática y reemplazar dP_Λ por δ_P , $d\phi_\Lambda$ por δ_ϕ y $d\phi_{\Lambda M}$ por δ_{ϕ_M}

$$\frac{E(\delta_P^2)}{P^2} = \frac{E(\delta_\phi^2)}{\phi^2} + \frac{E(\delta_{\phi_M}^2)}{\phi_M^2} - \frac{2E(\delta_\phi \delta_{\phi_M})}{\phi \phi_M}$$

Siendo $E(\delta_P^2)$ una estimación viciada de la varianza de la probabilidad de transvariación con sesgo igual a $\frac{n-1}{n}$ veces dicha varianza. Luego, la varianza insesgada de la probabilidad de transvariación es

$$\begin{aligned} D^2(P_\Lambda) &= \sigma_{P_\Lambda}^2 = \frac{n}{n-1} E(\delta_P^2) = \\ &= P^2 \left[\frac{D^2(\phi_\Lambda)}{\phi^2} + \frac{D^2(\phi_{M,\Lambda})}{\phi_M^2} - \frac{2\mu_{11}(\phi_\Lambda, \phi_{M,\Lambda})}{\phi \phi_M} \right] \end{aligned} \quad (8.60)$$

donde las varianzas $D^2(\phi_\Lambda)$ y $D^2(\phi_{M,\Lambda})$ y la covarianza $\mu_{11}(\phi_\Lambda, \phi_{M,\Lambda})$ están dadas, respectivamente, por (8.54), (8.55) y (8.56).

Para el caso particular de ser los parámetros λ iguales a las correspondientes medianas de las distribuciones consideradas y por lo tanto

$$p_M = q_M = \frac{1}{2} \quad (8.60.a)$$

resulta para la (8.60)

$$D^2(P_\Delta) = \frac{4}{n-1} \left(p q - p^2 - p^2 p_c - \frac{1}{4} p_0 \right) \quad (8.60.b)$$

8.9. — VARIANZA DE LA INTENSIDAD DE TRANSVARIACIÓN DE ORDEN r

La (8.27) nos define en el muestreo la v.a. intensidad de transvariación de orden r , correspondiente a la v.a. de (8.11). Elevando al cuadrado la diferencial logarítmica y tomando esperanza matemática, tenemos

$$\frac{E(\delta_{I_r}^2)}{I_r^2} = \frac{1}{r^2} \left[\frac{E(\delta_{m_r}^2)}{m_r^2} + \frac{E(\delta_{m_{r,M}}^2)}{m_{r,M}^2} - \frac{2 E(\delta_{m_r}, \delta_{m_{r,M}})}{m_r m_{r,M}} \right]$$

La $E(\delta_{I_r}^2)$ nos da una estimación sesgada o viciada de la varianza de la intensidad de transvariación de orden r . El sesgo, se demuestra, teniendo presente las conclusiones de los párrafos 8.6 y 8.7, es igual a $\frac{n-1}{n}$ veces dicha varianza. Luego, la varianza insesgada de la intensidad de transvariación de orden r es

$$D^2(I_{r,\Delta}) = \sigma_{I_{r,\Delta}}^2 = \frac{n}{n-1} E(\delta_{I_r}^2) = \quad (8.61)$$

$$= \frac{I_r^2}{r^2} \left[\frac{D^2(m_{r,\Delta})}{m_r^2} + \frac{D^2(m_{r,M,\Delta})}{m_{r,M}^2} - \frac{2 \mu_{11}(m_{r,\Delta}, m_{r,M,\Delta})}{m_r m_{r,M}} \right]$$

donde las varianzas $D^2(m_{r,\Delta})$ y $D^2(m_{r,M,\Delta})$ y la covarianza $\mu_{11}(m_{r,\Delta}, m_{r,M,\Delta})$ están dadas respectivamente por (8.57), (8.58) y (8.59).

8.10. — HIPOTESIS DE DISTRIBUCIÓN NORMAL

Los resultados precedentes son completamente generales, por lo que su estructura analítica deberá adaptarse en cada caso particular a las características propias de cada tipo de distribución.

En la hipótesis de distribución normal, es importante recordar que la *v.a.* así distribuida es una *v.a.* continua y que hemos trabajado con los parámetros λ iguales a los valores medianos.

En consecuencia tenemos, respectivamente para (8.54), (8.55) y (8.56), en la hipótesis de distribución normal, sin olvidar (8.60.a).

$$D^2(p_\Delta) = \frac{1}{n-1} p q = \frac{1}{n-1} \Phi(-d_1) \Phi(d_1) \quad (8.62)$$

$$D^2(p_{M,\Delta}) = \frac{1}{n-1} p_M q_M = \frac{1}{4(n-1)} \quad (8.63)$$

$$\mu_{11}(p_\Delta, p_{M,\Delta}) = \frac{1}{n-1} p q_M = \frac{1}{2(n-1)} \Phi(-d_1) \quad (8.64)$$

Con lo que la varianza de la probabilidad de transvariación dada en (8.60.b), resulta igual a

$$\begin{aligned} D^2(P_\Delta) &= \sigma_{P_\Delta}^2 = \frac{4}{n-1} p(q-p) = \\ &= \frac{4}{n-1} \Phi(-d_1) [\Phi(d_1) - \Phi(-d_1)] \end{aligned} \quad (8.65)$$

La varianza de la intensidad de transvariación de orden r , la obtenemos reemplazando en (8.61) los momentos m_r y $m_{r,M}$ que intervienen en su segundo miembro, como se desprende de (8.57), (8.58) y (8.59), por los resultados obtenidos, en la hipótesis de distribución normal en (3.43) y (3.45), respectivamente, donde los parámetros λ son iguales a las medias aritméticas de las distribuciones. Con respecto a $m_{r+r,M}$ que interviene en la covarianza, en (8.59), nos remitimos a (8.53.a) donde lo expresamos en función de las m_r .

8.II. — CASO DE MAS DE DOS DISTRIBUCIONES

Con lo precedentemente expuesto tenemos completado el estudio de la varianza de la probabilidad de transvariación, de la intensidad de transvariación de orden r y de sus respectivas componentes, correspondiente a la v.a. de (8.II). La estructura de las fórmulas obtenidas es completamente análoga y por lo tanto mantienen su validez para la v.a. reducida marginal k -dimensional introducida en (6.49) y para la v.a. reducida n -dimensional. Las componentes de (6.49), para cualquier hipótesis de distribución, están dadas en (6.32.a) y (6.32.b). En (6.31.a) y (6.31.b) tenemos las componentes de la v.a. reducida n -dimensional.

Queda entonces pendiente de estudio, la varianza de la probabilidad de transvariación y de la intensidad de transvariación de orden r , para más de dos distribuciones, que pasamos a considerar, en el caso general de v.a. reducidas marginales k -dimensionales.

8.II.I. — VARIANZA DE LA PROBABILIDAD DE TRANSVARIACION MARGINAL

Teniendo presente la hipótesis 8.I.d. y en el supuesto que las muestras de las s distribuciones sean de distinto tamaño, tenemos

a) para la varianza de la transvariabilidad marginal k -dimensional entre s distribuciones, dada por (7.12)

$$D^2 ({}_s p_{(k)\Delta}) = \sum_{i \neq j} \left[\frac{N_i N_j}{N(N - N_i)} \right]^2 D^2 (p_{ij,\Delta}) \quad (8.66)$$

donde $D^2 (p_{ij,\Delta})$ está dada por (8.54) y en la hipótesis de distribución normal por (8.62);

b) para la varianza del máximo de la transvariabilidad, dada por (7.13)

$$D^2 ({}_s p_{M(k)\Delta}) = \sum_{i \neq j} \left[\frac{N_i N_j}{N(N - N_i)} \right]^2 D^2 (p_{i,j;M,\Delta}) \quad (8.67)$$

donde la varianza del segundo miembro está dada por (8.55) y por (8.63) en la hipótesis de distribución normal ;

c) para la covarianza entre la transvariabilidad marginal k -dimensional y su máximo

$$\mu_{11}(\dot{s}P_{(k)\Delta}, \dot{s}P_{M(k)\Delta}) = \sum_{i \neq j} \left[\frac{N_i N_j}{N(N - N_i)} \right]^2 \mu_{11}(\dot{p}_{ij,\Delta}, \dot{p}_{ij,M,\Delta}) \quad (8.68)$$

donde la covarianza del segundo miembro está dada por (8.56) y en la hipótesis de distribución normal, por (8.64)

d) En forma semejante a la seguida para la obtención de (8.60) deducimos para la estimación insesgada de la varianza de la probabilidad de transvariación dada por (7.14)

$$D^2(\dot{s}P_{(k)\Delta}) = \dot{s}P_{(k)}^2 \left[\frac{D^2(\dot{s}P_{(k)\Delta})}{\dot{s}P_{(k)}^2} + \frac{D^2(\dot{s}P_{M(k)\Delta})}{\dot{s}P_{M(k)}^2} - \frac{2\mu_{11}(\dot{s}P_{(k)\Delta}, \dot{s}P_{M(k)\Delta})}{\dot{s}P_{(k)}\dot{s}P_{M(k)}} \right] \quad (8.69)$$

donde las varianzas del segundo miembro están dadas, respectivamente, por (8.66) y (8.67) y la covarianza por (8.68)

d₁) La varianza obtenida en (8.69), toma la siguiente forma, cuando es $N_1 = N_2 = \dots = N_s = \frac{N}{s}$, que corresponde a la (7.17)

$$D^2(\dot{s}P_{(k)\Delta}) = \dot{s}P_{(k)}^2 \left[\frac{\sum_{i < j} D^2(\dot{p}_{ij,\Delta})}{(\sum_{i < j} \dot{p}_{ij})^2} + \frac{\sum_{i < j} D^2(\dot{p}_{ij,M,\Delta})}{(\sum_{i < j} \dot{p}_{ij,M})^2} - \frac{2 \sum_{i < j} \mu_{11}(\dot{p}_{ij,\Delta}, \dot{p}_{ij,M,\Delta})}{\sum_{i < j} \dot{p}_{ij} \sum_{i < j} \dot{p}_{ij,M}} \right] \quad (8.69.a)$$

que se puede deducir directamente a partir de (7.17) y recordando la hipótesis 8.i.d, o reemplazando en (8.69) las expresiones de (7.15) (7.16) y los resultados obtenidos en este caso particular para (8.66) (8.67) y (8.68).

d₂) Para la forma abreviada de la probabilidad de transvariación dada en (7.21) calculada con respecto a los valores medianos de las distribuciones y para muestras de diferente tamaño, resultado que corresponde a la (7.18), la fórmula general de la varianza obtenida en (8.69) toma la siguiente forma

$$D^2(\dot{s}P_{(k)\Delta}) = 4 D^2(\dot{s}P_{(k)\Delta}) + 16 \dot{s}P_{(k)} [\dot{s}P_{(k)} D^2(\dot{s}P_{M(k)\Delta}) - \mu_{11}(\dot{s}P_{(k)\Delta}, \dot{s}P_{M(k)\Delta})] \quad (8.69.b)$$

siendo

$$D^2 ({}_s p_{(k)\Lambda}) = \sum_{i=1}^{s-1} \left[\frac{N_i \sum_{h=i+1}^s N_h}{N(N-N_i)} \right]^2 D^2 (p_{i(i+1,\dots,s)\Lambda}) + \quad (8.70)$$

$$+ \sum_{i=2}^s \left[\frac{N_i \sum_{h=1}^{i-1} N_h}{N(N-N_i)} \right]^2 D^2 (p_{i(1,2,\dots,i-1)\Lambda})$$

$$D^2 ({}_s p_{M(k)\Lambda}) = \sum_{i=1}^{s-1} \left[\frac{N_i \sum_{h=i+1}^s N_h}{N(N-N_i)} \right]^2 D^2 (p_{i(i+1,\dots,s)M,\Lambda}) + \quad (8.71)$$

$$+ \sum_{i=2}^s \left[\frac{N_i \sum_{h=1}^{i-1} N_h}{N(N-N_i)} \right]^2 D^2 (p_{i(1,2,\dots,i-1)M,\Lambda})$$

$$\mu_{11} ({}_s p_{(k)\Lambda}, {}_s p_{M(k)\Lambda}) = \sum_{i=1}^{s-1} \left[\frac{N_i \sum_{h=i+1}^s N_h}{N(N-N_i)} \right]^2 \mu_{11} (p_{i(i+1,\dots,s)\Lambda}, p_{i(i+1,\dots,s)M,\Lambda}) + \quad (8.72)$$

$$+ \sum_{i=1}^s \left[\frac{N_i \sum_{h=1}^{i-1} N_h}{N(N-N_i)} \right]^2 \mu_{11} (p_{i(1,2,\dots,i-1)\Lambda}, p_{i(1,2,\dots,i-1)M,\Lambda})$$

Las fórmulas correspondientes a las varianzas de los segundos miembros de (8.70) y (8.71) y a la covarianza del segundo miembro de (8.72), están dadas, respectivamente, por (8.54) (8.55) y (8.56), y por (8.62), (8.63) y (8.64) en la hipótesis de distribución normal, teniendo presente en ambos casos (8.60.a) y que n dependerá ahora del tamaño de las muestras correspondientes para cada i .

d_3) La varianza correspondiente a la (7.22) que nos da la probabilidad de transvariación marginal k -dimensional, en forma abreviada, para s muestras del mismo tamaño y trabajando con respecto a los valores medianos, es

$$D^2 ({}_s P_{(k)\Lambda}) = \frac{16}{s^2 (s-1)^2} \sum_{i=1}^{s-1} (s-i)^2 \cdot \quad (8.73)$$

$$\cdot [D^2 (p_{i(i+1,\dots,s)\Lambda}) + 4 {}_s p_{(k)}^2 D^2 (p_{i(i+1,\dots,s)M,\Lambda})] -$$

$$- \frac{64}{s^2 (s-1)^2} {}_s p_{(k)} \sum_{i=1}^{s-1} (s-i)^2 \mu_{11} (p_{i(i+1,\dots,s)\Lambda}, p_{i(i+1,\dots,s)M,\Lambda})$$

8.II.2. — VARIANZA DE LA INTENSIDAD DE TRANSVARIACIÓN MARGINAL.

Siguiendo un razonamiento paralelo al de 8.II.1 y siempre en la hipótesis 8.I.d, tenemos:

a) para la varianza del momento de transvariación marginal k -dimensional de orden r , entre las s distribuciones consideradas y para muestras de distinto tamaño, que corresponde a la (7.23), es

$$D^2 ({}_s m_{r(k), \Lambda}) = \sum_{i \neq j} \left[\frac{N_i N_j}{N(N - N_i)} \right]^2 D^2 (m_{ij; r, \Lambda}) \quad (8.74)$$

donde la varianza del segundo miembro está dada por (8.57) con o sin hipótesis de distribución normal;

b) para la varianza del máximo del momento de transvariación marginal dado por (7.24)

$$D^2 ({}_s m_{r, M(k), \Lambda}) = \sum_{i \neq j} \left[\frac{N_i N_j}{N(N - N_i)} \right]^2 D^2 (m_{ij; r, M, \Lambda}) \quad (8.75)$$

donde la varianza del segundo miembro está dada por (8.58)

c) para la covarianza entre las *v.a.* en el muestreo del momento de transvariación marginal k -dimensional de orden r y su máximo

$$\mu_{11} ({}_s m_{r(k), \Lambda}, {}_s m_{r, M(k), \Lambda}) = \sum_{i \neq j} \left[\frac{N_i N_j}{N(N - N_i)} \right]^2 \mu_{11} (m_{ij; r, \Lambda}, m_{ij; r, M, \Lambda}) \quad (8.76)$$

donde la covarianza del segundo miembro está dada por (8.59).

d) En forma semejante a la seguida para la deducción de (8.61) tenemos para la estimación insesgada de la varianza de la intensidad de transvariación marginal k -dimensional de orden r , dada por (7.25)

$$D^2 ({}_s I_{r(k), \Lambda}) = \frac{{}_s I_{r(k)}^2}{r^2} \cdot \left[\frac{D^2 ({}_s m_{r(k), \Lambda})}{{}_s m_{r(k)}^2} + \frac{D^2 ({}_s m_{r, M(k), \Lambda})}{{}_s m_{r, M(k)}^2} - \frac{2 \mu_{11} ({}_s m_{r(k), \Lambda}, {}_s m_{r, M(k), \Lambda})}{{}_s m_{r(k)} {}_s m_{r, M(k)}} \right] \quad (8.77)$$

donde las varianzas y la covarianza del segundo miembro están dadas respectivamente por (8.74), (8.75) y (8.76).

d_1) Cuando las s muestras correspondientes a las distribuciones sometidas a estudio son del mismo tamaño, la intensidad de transvariación marginal toma la forma de (7.26). Para este caso particular se deduce directamente o a partir de (8.77)

$$D^2 ({}_sI_{r(k),\Lambda}) = \frac{{}_sI_{r(k)}^2}{r^2}. \quad (8.77.a)$$

$$\left[\frac{\sum_{i < j} D^2 (m_{ij,r,\Lambda})}{(\sum_{i < j} m_{ij,r})^2} + \frac{\sum_{i < j} D^2 (m_{ij,r,M,\Lambda})}{(\sum_{i < j} m_{ij,r,M})^2} - \frac{2 \sum_{i < j} \mu_{11} (m_{ij,r,\Lambda}, m_{ij,r,M,\Lambda})}{\sum_{i < j} m_{ij,r} \sum_{i < j} m_{ij,r,M}} \right]$$

donde las varianzas del segundo miembro están dadas, respectivamente, por (8.57) y (8.58) y la covarianza por (8.59).

d_2) La varianza correspondiente a la forma abreviada de la intensidad de transvariación marginal k -dimensional de orden uno, dada por (7.30), se deduce que es igual a la obtenida en (8.77) para $r = 1$, siendo respectivamente, las varianza y la covarianza de su segundo miembro, en este caso particular

$$D^2 ({}_sm_{1(k),\Lambda}) = \sum_{i=1}^{s-1} \left[\frac{N_i \sum_{j=i+1}^s N_j}{N(N-N_i)} \right]^2 D^2 (m_{i(i+1,\dots,s)1,\Lambda}) +$$

$$+ \sum_{i=2}^s \left[\frac{N_i \sum_{j=1}^{i-1} N_j}{N(N-N_i)} \right]^2 D^2 (m_{i(1,\dots,i-1)1,\Lambda}) \quad (8.78)$$

$$D^2 ({}_sm_{r,M(k),\Lambda}) = \sum_{i=1}^{s-1} \left[\frac{N_i \sum_{j=i+1}^s N_j}{N(N-N_i)} \right]^2 D^2 (m_{i(i+1,\dots,s)1,M,\Lambda}) +$$

$$+ \sum_{i=2}^s \left[\frac{N_i \sum_{j=1}^{i-1} N_j}{N(N-N_i)} \right]^2 D^2 (m_{i(1,\dots,i-1)1,M,\Lambda}) \quad (8.79)$$

$$\begin{aligned}
& \mu_{11} ({}_s m_{r(k)\Lambda} , {}_s m_{r,M(k),\Lambda}) = \\
& = \sum_{i=1}^{s-1} \left[\frac{N_i \sum_{j=i+1}^s N_j}{N(N-N_i)} \right]^2 \mu_{11} (m_{i(i+1,\dots,s)1,\Lambda} , m_{i(i+1,\dots,s)1,M,\Lambda}) + \quad (8.80) \\
& + \sum_{i=2}^s \left[\frac{N_i \sum_{j=1}^{i-1} N_j}{N(N-N_i)} \right]^2 \mu_{11} (m_{i(1,\dots,i-1)1,\Lambda} , m_{i(1,\dots,i-1)1,M,\Lambda})
\end{aligned}$$

Las varianzas y la covarianza de los segundos miembros de (8.78) (8.79) y (8.80) están dadas respectivamente por (8.57), (8.58), (8.59), observando solo que la n de dichas fórmulas dependerá en cada caso, del tamaño de las muestras correspondientes a cada valor de i .

d_3) La varianza correspondiente a la (7.31) que nos da la intensidad de transvariación marginal k -dimensional de orden uno, para s muestras del mismo tamaño, calculada con respecto a la media aritmética, resulta

$$\begin{aligned}
D^2 ({}_s I_{1(k)\Lambda}) = & {}_s I_{1(k)}^2 \left[\frac{\sum_{i=1}^{s-1} (s-i)^2 D^2 (m_{i(i+1,\dots,s)1,\Lambda})}{\left(\sum_{i=1}^{s-1} (s-i) m_{i(i+1,\dots,s)1} \right)^2} + \right. \\
& + \frac{\sum_{i=1}^{s-1} (s-i)^2 D^2 (m_{i(i+1,\dots,s)1,M,\Lambda})}{\left(\sum_{i=1}^{s-1} (s-i) m_{i(i+1,\dots,s)1,M} \right)^2} - \quad (8.81) \\
& \left. \frac{2 \sum_{i=1}^{s-1} (s-i)^2 \mu_{11} (m_{i(i+1,\dots,s)1,\Lambda} , m_{i(i+1,\dots,s)1,M,\Lambda})}{\sum_{i=1}^{s-1} (s-i) m_{i(i+1,\dots,s)1} \sum_{i=1}^{s-1} (s-i) m_{i(i+1,\dots,s)1,M}} \right]
\end{aligned}$$

d_4) La varianza de la intensidad de transvariación marginal k -dimensional deducida precedentemente para distintos supuestos, se mantiene sin modificación en la hipótesis de distribución normal, en cuya aplicación solo debemos recurrir a (3.43) y (3.45), respectivamente, para el cálculo del momento de transvariación marginal k -dimensional de orden r y de su máximo, cuando trabajamos en esta hipótesis con respecto a la media.

9 - APLICACIONES

9.1. — INTRODUCCIÓN

A continuación nos ocupamos de las aplicaciones a la teoría de la transvariación desarrollada en los capítulos precedentes. Para ello hemos tomado series industriales de la economía argentina.

Las aplicaciones se realizan sobre muestras de ochenta y tres observaciones cada una, cuyos valores son los números índices clasificados por rama de industria ⁽¹⁰⁾, correspondientes a los años 1955, 1956 y 1957, referidos al año 1952 como año base. Dichos números índices son particulares determinaciones de las siguientes v.a.

- (9.1)
- 1) Índice de los obreros ocupados ;
 - 2) índice de las horas-obrero trabajadas ;
 - 3) índice del poder adquisitivo del monto de los salarios pagados ;
 - 4) índice del volumen físico de la producción industrial.

Al realizar para cada capítulo las aplicaciones propuestas, precisaremos la notación correspondiente a las muestras consideradas.

9.2. — TRANSVARIACIÓN EN R_2 - APLICACIONES AL CAPÍTULO 3

En el capítulo 3 nos ocupamos de la transvariación entre una v.a. bidimensional y un parámetro dado por un punto en el espacio de dos dimensiones.

La muestra considerada es la de los números índices correspondiente al año 1956, por lo que disponemos en consecuencia de cuatro muestras de ochenta y tres observaciones cada una, cuyos valores son particulares determinaciones de las v.a. dadas

⁽¹⁰⁾ DIRECCIÓN NACIONAL DE ESTADÍSTICA Y CENSOS — *Boletín Mensual de Estadística de la República Argentina* - Buenos Aires.

en 9.1. Luego, tendremos para estudiar en esta muestra tantos casos de transvariación en R_2 , como combinaciones posibles entre dichas cuatro v.a. hayan, es decir seis.

Simbolizamos respectivamente con ξ_i ($i = 1, 2, 3, 4$), cada una de las cuatro v.a. consideradas y con x_i ($i = 1, 2, 3, 4$), sus particulares determinaciones en el muestreo, es decir, sus correspondientes variables estadísticas.

Con \bar{x}_i y M_i ($i = 1, 2, 3, 4$) simbolizamos respectivamente la media y el valor mediano de las muestras consideradas, con lo que tenemos

$$\begin{array}{ll}
 (9.2) \quad \bar{x}_1 = 99,9 & M_1 = 99,7 \\
 \bar{x}_2 = 102,5 & M_2 = 100,0 \\
 \bar{x}_3 = 118,2 & M_3 = 113,0 \\
 \bar{x}_4 = 121,0 & M_4 = 118,5
 \end{array}$$

Teniendo presente (8.8), donde definimos c , en aplicaciones con valores discretos, como el entero más próximo, por defecto o por exceso a $\frac{\lambda_1 - \lambda_2}{h}$, resulta, según trabajemos con λ igual a la media o a la mediana, los siguientes valores :

(9.3)

VARIABLES ESTADÍSTICAS x_i con x_j		h (1)	$\overset{c}{c}$ (medias) (2)	$\overset{c}{c}$ (medianas) (3)
i	j			
1	2	3	1	0
1	3	3	6	4
1	4	3	7	6
2	3	3	5	4
2	4	3	6	6
3	4	3	1	2

Para la estimación en el muestreo de los parámetros poblacionales de la probabilidad e intensidad de transvariación, con sus respectivas componentes, en nuestro e.a., aplicamos (8.22) a (8.27), mientras no adoptemos ninguna hipótesis de v.a. continua, sobre la forma de la distribución madre.

Para la particular determinación en el muestreo de la v.a. aquí considerada, es decir su correspondiente variable estadística, nos remitimos a (8.12).

Para la probabilidad de transvariación trabajamos en las aplicaciones, sin excepción, con los valores medianos de las v.a. consideradas. En consecuencia resulta :

$$(9.4) \quad p_M = q_M = \frac{1}{2}$$

por las razones dadas en los capítulos precedentes.

9.2.a — PROBABILIDAD DE TRANSVARIACIÓN

A partir de la Tabla A que nos da los valores estadísticos de las 83 ramas de industria consideradas, para el año 1956, correspondiente a las cuatro v.a. indicadas en (9.1), elaboramos los cuadros N° 1 a N° 6 que corresponden respectivamente a las seis combinaciones binarias posibles entre dichas cuatro v.a.

Llevando a (8.22) y (8.26) los valores de los cuadros N° 1 a N° 6 y teniendo presente (9.3), resulta la Tabla N° 1.

La columna (4) nos da el valor de la transvariabilidad correspondiente a cada una de las combinaciones binarias consideradas. La columna (5) nos da el valor de la probabilidad de transvariación.

A partir de (8.54), (8.55) y (8.56) calculamos, respectivamente, la varianza de la transvariabilidad, del máximo de la transvariabilidad y la covarianza entre estas dos v.a. en el muestreo. A partir de (8.60) o (8.60.b), calculamos la varianza de la probabilidad de transvariación. Los resultados los consignamos en la Tabla N° 2.

Observamos que no obstante trabajar con el valor mediano de las distribuciones calculamos la varianza del máximo de la

transvariabilidad, debido a que el valor mediano del colectivo lo estimamos a través de las muestras y por lo tanto en ella es una nueva v.a. cuya varianza está dada para cada caso por la columna (2).

Siendo los valores de la columna (5) de la Tabla N° 1 una estimación insesgada o eficiente de la probabilidad de transvariación en el colectivo, resulta que en el muestreo, en virtud de las conclusiones del teorema central del límite, se distribuye normalmente, con media igual a su esperanza matemática y con varianza dada por los correspondientes valores de la columna (4) en la Tabla N° 2.

Fijando en el 5 % el nivel de confianza, la probabilidad de transvariación en el colectivo estará comprendida, con un coeficiente de confianza igual a 0,95, en el intervalo

$$\begin{array}{lll}
 (9.5) & 0,633 < P_{12} < 0,837 & \text{para } x_1 \text{ con } x_2 \\
 & 0,027 < P_{13} < 0,214 & \text{» } x_1 \text{ » } x_3 \\
 & 0,215 < P_{14} < 0,484 & \text{» } x_1 \text{ » } x_4 \\
 & 0,013 < P_{23} < 0,180 & \text{» } x_2 \text{ » } x_3 \\
 & 0,331 < P_{24} < 0,609 & \text{» } x_2 \text{ » } x_4 \\
 & 0,855 < P_{34} < 0,977 & \text{» } x_3 \text{ » } x_4
 \end{array}$$

9.2.b — INTENSIDAD DE TRANSVARIACIÓN DE ORDEN UNO

Con la información provista por los cuadros N° 1 a N° 6, obtenida a partir de la Tabla A, construimos la Tabla N° 3.

Llevando a (8.24), (8.25) y (8.27) los valores de la Tabla N° 3, obtenemos la Tabla N° 4, que nos da respectivamente, los momentos de transvariación de orden uno y dos, los máximos de estos momentos y la intensidad de transvariación de orden uno.

Llevando los valores de la Tabla N° 4 a (8.57), (8.58), (8.59) y (8.61) construimos la Tabla N° 5, que nos da respectivamente, la varianza del momento de transvariación de orden uno, la varianza de su máximo, la covarianza entre estos dos momentos muestrales y la varianza de la intensidad de transvariación de

orden uno. En la última columna consignamos el desvío medio cuadrático de la intensidad de transvariación.

La estimación muestral de la columna (5) de la Tabla N° 4, por lo demostrado en 8.4, es una estimación insesgada o eficiente de la intensidad de transvariación poblacional. Para grandes muestras, en virtud de las conclusiones del teorema central del límite, la intensidad de transvariación muestral se distribuye normalmente, es decir que converge en probabilidad a la distribución normal, con media igual a la media poblacional y varianza dada en nuestro caso particular por los valores de la columna (4) de la Tabla N° 5.

Fijando en el 5 % el nivel de confianza, la intensidad de transvariación poblacional estará comprendida, con un coeficiente de confianza igual al 95 %, en el intervalo

$$\begin{array}{ll}
 (9.6) & 0,552 < I_{12;1} < 0,711 \quad \text{para } x_1 \text{ con } x_2 \\
 & 0,026 < I_{13;1} < 0,033 \quad \quad \quad \text{'' } x_1 \text{ '' } x_3 \\
 & 0,049 < I_{14;1} < 0,205 \quad \quad \quad \text{'' } x_1 \text{ '' } x_4 \\
 & 0,006 < I_{23;1} < 0,050 \quad \quad \quad \text{'' } x_2 \text{ '' } x_3 \\
 & 0,095 < I_{24;1} < 0,288 \quad \quad \quad \text{'' } x_2 \text{ '' } x_4 \\
 & 0,844 < I_{34;1} < 0,901 \quad \quad \quad \text{'' } x_3 \text{ '' } x_4
 \end{array}$$

9.2.c — HIPOTESIS DE DISTRIBUCIÓN NORMAL

A partir de los cuadros N° 1 a N° 6, obtenemos los valores de la Tabla N° 6.

Las columnas (1) y (2) nos dan una estimación muestral insesgada de las varianzas de las distribuciones marginales consideradas; las columnas (3) y (4), el coeficiente de correlación y la varianza de la diferencia de las v.a. componentes de las correspondientes distribuciones bidimensionales. En cuanto a la columna (6), ella se obtiene llevando a (3.41) los valores de la columna (4) y (5).

En la hipótesis de distribución normal, las seis distribuciones se encuentran normalmente correlacionadas. En efecto, aplicán-

doles a los coeficientes de correlación la hipótesis nula de R.A. Fisher, resulta la v.a. transformada.

$$(9.7) \quad t = \frac{r}{\sqrt{1-r^2}} \sqrt{n-2}$$

que tiene una distribución de Student, con $n-2$ grados de libertad, es decir que en el muestreo, tiene dicha distribución exacta.

Siendo para todos los casos $n = 83$, demostramos a continuación que el menor de los coeficientes de correlación, o sea r_{34} , correspondiente a las variables estadísticas x_3 con x_4 , no puede provenir de un colectivo incorrelacionado, para un nivel de confianza prefijado, con lo que queda demostrado para todos.

Para $r_{34} = 0,4965$, $n = 83$ y un nivel de confianza igual al 5% resulta:

$$(9.8) \quad t = \frac{4,4685}{0,868} = 5,148$$

$$(9.9) \quad P(|t| \geq 5,148) = 0,000.002.64$$

resultado altamente significativo que no puede provenir de un colectivo sin correlacionar, con lo que la hipótesis nula queda rechazada para las seis combinaciones binarias.

9.2.c₁ — PROBABILIDAD DE TRANSVARIACIÓN

A partir de las estimaciones de los parámetros poblacionales consignadas en la Tabla N° 6 y teniendo presente (3.56) y (8.65), deducimos las correspondientes estimaciones muestrales de la transvariabilidad, probabilidad de transvariación y varianza de la probabilidad de transvariación para las distribuciones bidimensionales consideradas, que juntamente con el desvío medio cuadrático consignamos en la Tabla N° 7.

Los valores de la columna (2) de la Tabla N° 7, son estimaciones insesgadas o eficientes de las probabilidades de transvariación poblacionales, por lo demostrado en el párrafo 8.4. En el mues-

treo, dicha estimación se distribuye normalmente, en virtud de las conclusiones del teorema central del límite, con media igual a su esperanza matemática y varianza dada por los valores de la columna (4).

Fijando en el 5 % el nivel de confianza, la probabilidad de transvariación poblacional estará comprendida, en la hipótesis de normalidad y con un coeficiente de confianza igual al 95 %, en el intervalo

$$\begin{array}{lll}
 (9.10) & 0,634 < P_{12} < 0,894 & \text{para } x_1 \text{ con } x_2 \\
 & 0,059 < P_{13} < 0,291 & \text{» } x_1 \text{ » } x_3 \\
 & 0,284 < P_{14} < 0,588 & \text{» } x_1 \text{ » } x_4 \\
 & 0,075 < P_{23} < 0,318 & \text{» } x_2 \text{ » } x_3 \\
 & 0,354 < P_{24} < 0,660 & \text{» } x_2 \text{ » } x_4 \\
 & 0,846 < P_{34} \leq 1,000 & \text{» } x_3 \text{ » } x_4
 \end{array}$$

9.2.c₂ — INTENSIDAD DE TRANSVARIACIÓN

A partir de las estimaciones de los parámetros poblacionales de las distribuciones sometidas a estudio, consignadas en la Tabla N° 6 y recordando (3.53), (3.54) y (3.58), deducimos los momentos de transvariación de orden uno y dos, los máximos de dichos momentos y la intensidad de transvariación de orden uno, que detallamos en la Tabla N° 8.

Llevando los valores de la Tabla N° 8 a (8.57), (8.58), (8.59) y (8.61), tenemos respectivamente, para $r = 1$, la varianza del momento de transvariación de orden uno, de su máximo, la covarianza entre ambos momentos y la varianza de la intensidad de transvariación de orden uno, que se detallan en la Tabla N° 9.

La estimación muestral de la columna (5) de la Tabla N° 8, para grandes muestras, converge en probabilidad a la distribución normal, con media igual a su esperanza matemática, por ser una estimación insesgada o eficiente y con varianza dada en nuestro caso particular, por los valores de la columna (4) de la Tabla N° 9.

Fijando en el 5 % el nivel de confianza, la intensidad de transvariación de orden uno del colectivo, estará comprendida, con un coeficiente de confianza igual al 95 %, en el intervalo

$$\begin{array}{lll}
 (9.11) & 0,603 < I_{12;1} < 0,734 & \text{para } x_1 \text{ con } x_2 \\
 & 0,024 < I_{13;1} < 0,178 & \text{» } x_1 \text{ » } x_3 \\
 & 0,214 < I_{14;1} < 0,412 & \text{» } x_1 \text{ » } x_4 \\
 & 0,035 < I_{23;1} < 0,197 & \text{» } x_2 \text{ » } x_3 \\
 & 0,283 < I_{24;1} < 0,478 & \text{» } x_2 \text{ » } x_4 \\
 & 0,787 < I_{34;1} < 0,988 & \text{» } x_3 \text{ » } x_4
 \end{array}$$

9.3. — TRANSVARIACIÓN ENTRE DOS DISTRIBUCIONES EN R_n — APLICACIÓN AL CAPÍTULO 6.

En las aplicaciones al capítulo 6, tomamos dos muestras de cuatro dimensiones. Las componentes de cada una de estas dos distribuciones corresponden a los conceptos indicados en (9.1) y los valores muestrales están consignados en las Tablas B y C, que corresponden a los índices industriales del año 1955 y 1957 respectivamente, siendo en todos los casos de tamaño 83. En consecuencia resultan quince los casos posibles de transvariación: 1 de transvariación entre dos distribuciones de cuatro dimensiones; 4 transvariaciones marginales tridimensionales; 6 bidimensionales y 4 unidimensionales.

Simbolizamos como en (6.31.a) y (6.31.b) nuestras v.a. reducidas de cuatro dimensiones y con τ_i y ν_i ($i = 1, 2, 3, 4, 12, \dots, 234$), las correspondientes v.a. reducidas marginales k -dimensionales ($k = 1, 2$, y 3). Con x e y simbolizamos las variables estadísticas correspondientes a τ y ν , y con x_i e y_i ($i = 1, 2, \dots, 234$), las variables estadísticas correspondientes a τ_i y ν_i .

A partir de las Tablas B y C, resulta:

$$\begin{array}{ll}
 (9.12) & \Delta_{11} = 20,64 \\
 & \Delta_{22} = 22,68 \\
 & \Delta_{33} = 31,40 \\
 & \Delta_{44} = 43,90
 \end{array}$$

que como divisores de las tablas B y C, nos dan por resultados los valores reducidos de la tabla D. A partir de la Tabla D, formamos los cuadros N° 9 a N° 23, correspondientes a cada una de las trece combinaciones posibles de las v.a. reducidas que se consideran.

Con los valores de los cuadros N° 9 a N° 23, obtenemos la Tabla N° 10.

Las variables reducidas correspondientes a los cuadros N° 9 a 23, reciben un tratamiento completamente equivalente al aplicado a las variables estadísticas de los cuadros N° 1 a N° 6, como se desprende del contenido de 6.7. Manteniendo las mismas consideraciones y sistema de hipótesis dado en 9.2, pasamos a aplicar las conclusiones de los capítulos 6 y 8 a los e.a. aquí considerados.

9.3.a — PROBABILIDAD DE TRANSVARIACIÓN MARGINAL

Llevando los valores obtenidos en los cuadros N° 9 a N° 23 a (8.22), (8.26), (8.54), (8.55), (8.56) y (8.60) o (8.60.b), se llega a los resultados que se consignan en la tabla N° 11.

Para un nivel de confianza igual al 5 %, la probabilidad de transvariación poblacional estará comprendida, con un coeficiente de confianza igual al 95 %, en el intervalo,

(9.13)	$0,849 < P_1 < 0,910$	para x_1 con y_1
	$0,931 < P_2 < 0,997$	» x_2 » y_2
	$0,396 < P_3 < 0,640$	» x_3 » y_3
	$0,611 < P_4 < 0,714$	» x_4 » y_4
	$0,083 < P_{12} < 0,303$	» x_{12} » y_{12}
	$0 \leq P_{13} < 0,071$	» x_{13} » y_{13}
	$0 \leq P_{14} < 0,071$	» x_{14} » y_{14}
	$0 \leq P_{23} < \varepsilon$	» x_{23} » y_{23}
	$0 \leq P_{24} < \varepsilon$	» x_{24} » y_{24}
	$0 \leq P_{34} < \varepsilon$	» x_{34} » y_{34}
	$0 \leq P_{123} < \varepsilon$	» x_{123} » y_{123}

$$\begin{array}{ll}
0 \leq P_{124} < \varepsilon & \text{para } x_{124} \text{ con } y_{124} \\
0 \leq P_{134} < 0,071 & \text{» } x_{134} \text{ » } y_{134} \\
0 \leq P_{234} < \varepsilon & \text{» } x_{234} \text{ » } y_{234} \\
0 \leq P_{1234} < \varepsilon & \text{» } x_{1234} \text{ » } y_{1234}
\end{array}$$

Cuando para un e.a. se obtiene, como en algunos de los casos precedentes, una probabilidad nula, ello no significa necesariamente, en virtud de la interpretación frecuencial de la probabilidad, que el suceso esperado sea de presentación imposible. Solo implica que, para grandes muestras, se obtendrán valores de P en el semientorno positivo de cero, es decir que reiterando la realización del e.a. un número suficientemente grande de veces, en condiciones homogéneas, el suceso esperado podrá presentarse, a lo sumo en un por ciento muy reducido. Esta conclusión se extiende válidamente para el caso más general de ser $0 \leq P < \varepsilon$ para ε positivo y suficientemente pequeño. La probabilidad de presentación de un suceso cuya probabilidad está comprendida en este intervalo, es prácticamente nula, en una sola realización del e.a. considerado.

A conclusiones opuestas se llega para valores de P en el semientorno lateral izquierdo de uno. Si en la realización del e.a. resulta $P = 1$, ello no implica necesariamente que el suceso esperado sea de presentación necesaria, sino que el mismo se presentará en un por ciento muy elevado, reiterando un número suficientemente grande de veces el e.a. considerado, en condiciones homogéneas. Extendemos esta conclusión al semientorno de la unidad

$$1 - \varepsilon < P \leq 1$$

siendo prácticamente segura la presentación del suceso esperado en una sola realización del e.a. dado.

9.3.b — INTENSIDAD DE TRANSVARIACION MARGINAL

Con los valores provistos por los cuadros N° 9 a N° 23, la Tabla N° 10 y las fórmulas dadas por (8.24), (8.25), (8.27), (8.57) (8.58), (8.59) y (8.61), obtenemos los resultados que se consignan en la Tabla N° 12 y N° 13.

Para un nivel de confianza igual al 5 %, la intensidad de transvariación poblacional de orden uno, estará comprendida, con un coeficiente de confianza igual al 95 %, en el intervalo :

(9.14)	$1 - \varepsilon < I_{1,1} \leq 1$	para x_1	con y_1
	$0,998 < I_{2,1} \leq 1$	» x_2	» y_2
	$0,045 < I_{3,1} < 0,131$	» x_3	» y_3
	$0,119 < I_{4,1} < 0,333$	» x_4	» y_4
	$0 \leq I_{12,1} < 0,046$	» x_{12}	» y_{12}
	$0 \leq I_{13,1} < 0,003$	» x_{13}	» y_{13}
	$0 \leq I_{14,1} < 0,035$	» x_{14}	» y_{14}
	$0 \leq I_{23,1} < \varepsilon$	» x_{23}	» y_{23}
	$0 \leq I_{24,1} < \varepsilon$	» x_{24}	» y_{24}
	$0 \leq I_{34,1} < \varepsilon$	» x_{34}	» y_{34}
	$0 \leq I_{123,1} < \varepsilon$	» x_{123}	» y_{123}
	$0 \leq I_{124,1} < \varepsilon$	» x_{124}	» y_{124}
	$0,002 < I_{134,1} < 0,003$	» x_{134}	» y_{134}
	$0 \leq I_{234,1} < \varepsilon$	» x_{234}	» y_{234}
	$0 \leq I_{1234,1} < \varepsilon$	» x_{1234}	» y_{1234}

Las intensidades de transvariación que asumen valores en el semientorno lateral derecho de cero y en el semientorno lateral izquierdo de la unidad, con semiamplitud ε , reciben análogas interpretaciones a las realizadas para las probabilidades de transvariación en 9.3.a.

9.3.c — HIPOTESIS DE DISTRIBUCION NORMAL

A partir de los cuadros N° 9 a N° 23, construimos la Tabla N° 14.

Estas distribuciones, en la hipótesis normal, se encuentran normalmente correlacionadas, pues al aplicar la hipótesis nula a los coeficientes de correlación, se demuestra, como en 9.2.c, que dichas muestras no pueden provenir, para cualquier nivel

de confianza usual, de un colectivo sin correlacionar. Al rechazar la hipótesis nula, pierde todo interés el cálculo del área o espacio de transvariación y el valor discriminativo, conceptos introducidos para los casos de v.a. independientes.

9.3.c₁ — PROBABILIDAD DE TRANSVARIACIÓN MARGINAL

Llevando los valores de la Tabla N° 14 a (3.56) y (8.65), obtenemos la Tabla N° 15.

Para un nivel de confianza igual al 5 %, la probabilidad de transvariación poblacional, estará comprendida, con un coeficiente de confianza igual al 95 %, en el intervalo

(9.15)	$0,787 < P_1 < 0,983$	para x_1	con y_1
	$1 - \varepsilon < P_2 \leq 1$	» x_2	» y_2
	$0,480 < P_3 < 0,776$	» x_3	» y_3
	$0,643 < P_4 < 0,900$	» x_4	» y_4
	$0,194 < P_{12} < 0,484$	» x_{12}	» y_{12}
	$0 \leq P_{13} < 0,042$	» x_{13}	» y_{13}
	$0 \leq P_{14} < 0,118$	» x_{14}	» y_{14}
	$0 \leq P_{23} < 0,000.154$	» x_{23}	» y_{23}
	$0 \leq P_{24} < 0,000.611$	» x_{24}	» y_{24}
	$0 \leq P_{34} < 0,000.463$	» x_{34}	» y_{34}
	$0 \leq P_{123} < 0,002$	» x_{123}	» y_{123}
	$0 \leq P_{124} < 0,046$	» x_{124}	» y_{124}
	$0 \leq P_{134} < 0,154$	» x_{134}	» y_{134}
	$0 \leq P_{234} < 0,000.283$	» x_{234}	» y_{234}
	$0 \leq P_{1234} < 0,002$	» x_{1234}	» y_{1234}

9.3.c₂ — INTENSIDAD DE TRANSVARIACION MARGINAL

Llevando los valores de la Tabla N° 14 a (3.53), (3.54), (3.58), (8.57), (8.58), (8.59) y (8.61), obtenemos los resultados que se consignan en las Tablas N° 16 y N° 17.

Para un nivel de confianza igual al 5 %, la intensidad de transvariación marginal poblacional de orden uno, estará comprendida, con un coeficiente de confianza igual al 95 %, en el intervalo.

(9.16)	$0,792 < I_{1,1} < 0,867$	para x_1	con y_1
	$1-\varepsilon < I_{2,1} \leq 1$	» x_2	» y_2
	$0,421 < I_{3,1} < 0,595$	» x_3	» y_3
	$0,614 < I_{4,1} < 0,742$	» x_4	» y_4
	$0,130 < I_{12,1} < 0,324$	» x_{12}	» y_{12}
	$0 \leq I_{13,1} < 0,021$	» x_{13}	» y_{13}
	$0 \leq I_{14,1} < 0,065$	» y_{14}	» y_{14}
	$0 \leq I_{23,1} < 0,000.052$	» x_{23}	» y_{23}
	$0 \leq I_{24,1} < 0,000.216$	» x_{24}	» y_{24}
	$0 \leq I_{34,1} < 0,000.521$	» x_{34}	» y_{34}
	$0 \leq I_{123,1} < 0,000.297$	» x_{123}	» y_{123}
	$0 \leq I_{124,1} < 0,012$	» x_{124}	» y_{124}
	$0 \leq I_{134,1} < 0,087$	» x_{134}	» y_{134}
	$0 \leq I_{234,1} < 0,000.093$	» x_{234}	» y_{234}
	$0 \leq I_{1234,1} < 0,001.301$	» x_{1234}	» y_{1234}

9.4. — TRANSVARIACION ENTRE MAS DE DOS DISTRIBUCIONES n-DIMENSIONALES-APLICACIONES AL CAPITULO 7

Para las aplicaciones a este capítulo, tomamos cuatro muestras unidimensionales de tamaño 83, para el año 1956, cuyas v.a. corresponden a los conceptos indicados en (9.1) y sus correspondientes valores muestrales se consignan en la Tabla A.

Simbolizamos como en 9.2, con ξ_i ($i = 1, 2, 3, 4$) cada una de las cuatro v.a. consideradas y con x_i ($i = 1, 2, 3, 4$) sus correspondientes variables estadísticas.

La media y el valor mediano de los valores consignados en la Tabla A, están dados en (9.2).

En virtud de las relaciones de desigualdad que se observan tanto para las medias como para los valores medianos y a los

efectos de aplicar las fórmulas abreviadas de la probabilidad e intensidad de transvariación dadas en (7.22) y (7.31), por tratarse de muestras del mismo tamaño, formamos con las cuatro distribuciones consideradas los siguientes agrupamientos:

1) Índice del poder adquisitivo del monto de los salarios pagados con índice del volumen físico de la producción industrial;

2) Índice de las horas-obra trabajadas con la distribución aditiva de frecuencias de las dos distribuciones que intervienen en el punto anterior;

3) Índice de los obreros ocupados con la distribución aditiva de frecuencias de las tres distribuciones restantes.

Las distribuciones de frecuencia resultantes, se dan respectivamente, en los cuadros N° 6, N° 7 y N° 8. A partir de estos cuadros obtenemos los resultados de la Tabla N° 18.

Manteniendo el sistema de hipótesis dado en 9.2, pasamos a aplicar las conclusiones de los capítulos 7 y 8 al e.a. aquí consignado.

9.4.a — PROBABILIDAD DE TRANSVARIACIÓN

A partir de la información suministrada por los cuadros N° 6, N° 7 y N° 8 y las fórmulas dadas en (8.22), (8.26), (8.54), (8.55), (8.56) y (8.60) o (8.60.b), obtenemos los resultados de la Tabla N° 19.

Llevando estos resultados a (7.22), tenemos para la probabilidad de transvariación entre las cuatro distribuciones

$$(9.17) \quad {}_4P_{(1)} = 0,443.769$$

y teniendo presente que estamos trabajando con los valores medianos, resulta para la transvariabilidad

$$(9.18) \quad {}_4p_1 = 0,221.884.5$$

Llevando los resultados precedentes a (8.73), tenemos para la varianza y el desvío medio cuadrático de la probabilidad de transvariación

$$(9.19) \quad D^2({}_4P_{(1)\Delta}) = 0,000.752$$

$$(9.20) \quad D({}_4P_{(1)\Delta}) = 0,027.423$$

De donde resulta, para un nivel de confianza igual al 5 %

$$(9.21) \quad 0,390 < {}_4P_{(1)} < 0,498$$

9.4.b — INTENSIDAD DE TRANSVARIACIÓN

Con la información suministrada por los cuadros N° 6, N° 7 y N° 8 y las fórmulas dadas por (8.24), (8.25), (8.57), (8.58) y (8.59), elaboramos las Tablas N° 20 y N° 21. Llevando los resultados de estas tablas a (7.31) y (8.81), tenemos respectivamente la intensidad de transvariación de orden uno entre cuatro distribuciones y su varianza, a saber

$$(9.22) \quad {}_4I_1 = 0,211.691$$

$$(9.23) \quad D^2 ({}_4I_{1,\Delta}) = 0,000.453$$

Siendo el desvío medio cuadrático

$$(9.24) \quad D ({}_4I_{1,\Delta}) = 0,021.284$$

Con lo que tenemos, para un nivel de confianza igual al 5 %

$$(9.25) \quad 0,170 < {}_4I_1 < 0,253$$

9.4.c — HIPOTESIS DE DISTRIBUCION NORMAL

Con la información suministrada por los cuadros N° 6, N° 7 y N° 8, obtenemos los resultados de la Tabla N° 22.

Las distribuciones aquí consideradas, en la hipótesis de distribución normal, se encuentran normalmente correlacionadas, pues la hipótesis nula resulta rechazada, aún para altos coeficientes de confianza.

9.4.c₁ — PROBABILIDAD DE TRANSVARIACION

Llevando los valores de la Tabla N° 22 a (7.46.a), (3.56), (8.62), (8.63) y (8.64), obtenemos la Tabla N° 23. Con estos re-

sultados y las fórmulas dadas por (7.22) y (8.73), tenemos respectivamente

$$(9.26) \quad {}_4P_{(1)} = 0,530.280$$

$$(9.27) \quad {}_4\dot{P}_{(1)} = 0,265.140$$

$$(9.28) \quad D^2({}_4P_{(1)\Delta}) = 0,000.943$$

$$(9.29) \quad D({}_4P_{(1)\Delta}) = 0,030.708$$

Luego, para un nivel de confianza igual al 5 %, resulta

$$(9.30) \quad 0,470 < {}_4P_{(1)} < 0,590$$

9.4.c₂ — INTENSIDAD DE TRANSVARIACION

Con los valores de la Tabla N° 22 y las fórmulas (7.49), (7.50), (7.51.a), (7.51.b), (8.57), (8.58) y (8.59), construimos las Tablas N° 24 y N° 25. Llevando estos resultados a (7.31) y (8.81), tenemos

$${}_4I_1 = 0,454.970$$

$$D^2({}_4I_{1,\Delta}) = 0,000.608$$

$$D({}_4I_{1,\Delta}) = 0,024.658$$

De donde resulta, para un nivel de confianza igual al 5 %

$$0,407 < {}_4I_1 < 0,503$$

9.5. — CONCLUSIÓN

Recordando las características de la probabilidad e intensidad de transvariación, el significado de su definición y los resultados obtenidos en los e.a. considerados, resultan las siguientes conclusiones :

1) Los valores muestrales correspondientes al año 1956, expresados bajo la forma de números índices, con base 100 en el

año 1952, presentan un alto grado de probabilidad e intensidad de transvariación entre el índice de los obreros ocupados y el de las horas-obrero trabajadas y entre el índice del poder adquisitivo de los salarios pagados y el del volumen físico de la producción industrial, lo que revela un comportamiento semejante y en la misma dirección de cada uno de estos pares de índices, durante el período 1952-1956; ello significa, que si seleccionamos al azar una rama de industria en nuestro país y observamos el valor de sus cuatro índices dados por (9.1) para el año 1956, podemos esperar, con un alto grado de probabilidad, que los dos primeros índices y el de los dos últimos presenten, respectivamente, una diferencia de signo opuesto al signo de la diferencia de sus correspondientes medias o medianas poblacionales, o en otras palabras, sus medias y valores medianos, tienden a mantenerse muy próximos entre sí. No ocurre lo mismo con las cuatro combinaciones binarias restantes, es decir entre el índice de los obreros ocupados o el de las horas-obrero trabajadas, con el índice del poder adquisitivo del monto de los salarios pagados o el del volumen físico de la producción industrial, con lo que, en la selección al azar considerada, podemos esperar, con un alto grado de probabilidad, que estas cuatro combinaciones no presenten una diferencia de índices de signo opuesto al signo de la diferencia de sus medias o medianas poblacionales; ello significa que entre 1952 y 1956 se ha operado un distanciamiento entre las medias y medianas poblacionales de cada uno de los dos primeros índices de (9.1) con respecto a cada uno de los dos últimos, en virtud de la relación de desigualdad existente entre dichos valores medios; o sea que, mientras los índices de los obreros ocupados y de las horas-obrero trabajadas en 1956 han permanecido, en promedio, muy aproximadamente al nivel de 1952, lo que cabía esperar si admitimos el supuesto de la plena ocupación, los otros dos tuvieron un apreciable crecimiento, de donde surge la muy importante conclusión, opuesta por cierto a la opinión de los que emiten juicios en base a apreciaciones personales, que el aumento del poder adquisitivo del monto de los salarios pagados en la industria en 1956, con relación a 1952, representa un mejoramiento del nivel de vida del trabajador industrial, lo que unido al más importante crecimiento promedio del volumen físico de la producción, se desprende que en 1956 la productividad industrial, en relación al

factor trabajo, fué superior a 1952, dada la relativa constancia de los obreros ocupados y de las horas-obrero trabajadas. Esta mayor productividad, contribuye también, al redistribuirse entre los titulares de los factores de la producción, al mejoramiento señalado para el nivel de vida del trabajador industrial, sin que dicho mejoramiento logre alcanzar la magnitud del incremento operado en la producción, en el supuesto que el factor capital o el factor empresario no haya contribuído también a la mayor producción.

Pero estas iniciales conclusiones favorables en relación al comportamiento en el tiempo del nivel de vida y de la producción industrial, no agotan el problema, desde que el mismo no puede desvincularse de la capitalización del país y de su fuerza demográfica, en torno a los cuales giran aspectos fundamentales del desarrollo económico, como lo son la tasa de crecimiento demográfico y la tasa de capitalización ;

2) En cuanto al estudio de la transvariación entre los valores muestrales correspondientes a los años 1955 y 1957, sin olvidar las relaciones de desigualdad entre sus valores medios, resulta :

a) los índices de los obreros ocupados y de las horas-obrero trabajadas presentan en promedio una fuerte constancia de 1955 a 1957, como se desprenden de las estimaciones de sus parámetros de la transvariación, que son iguales o aproximadamente iguales a la unidad ;

b) los índices del poder adquisitivo de los salarios pagados y del volumen físico de la producción, presentan una relativamente alta probabilidad de transvariación y una baja intensidad de transvariación, discordancia debida principalmente a la alta frecuencia a_0 , en relación al tamaño de la muestra, la que no actúa en la intensidad de transvariación, por ser nula, en el punto $t = 0$, la medida de la transvariación. Estos resultados permiten adelantar la conclusión que el año 1957 registra un substancial aumento del poder adquisitivo de los salarios pagados y del volumen físico de la producción industrial, con respecto al año 1955 ;

c) la estimación de los parámetros de la transvariación marginal k -dimensional ($k = 2,3$) y de la transvariación n -dimensional ($n = 4$) presentan un valor nulo o próximo a cero, lo cual significa, recordando lo observado en los últimos párrafos

de 9.3.a, que es prácticamente nula la probabilidad de que, en una selección al azar de una rama de industria, sus valores reducidos de dos, tres y cuatro dimensiones, para los años 1955 y 1957, presenten una diferencia de signo opuesto al signo de la diferencia de sus correspondientes valores reducidos medios de la población a que pertenecen, es decir que las combinaciones binarias, ternarias y cuaternarias de los valores reducidos, teniendo presente la relación de desigualdad entre sus valores medios, se han distanciado apreciablemente, en promedio, en el año 1957, con relación a 1955. Esto equivale a expresar que si estamos en presencia de los valores reducidos de una rama de industria para los años 1955 y 1957, de los cuales solo sabemos que unos pertenecen a un año y otros al otro, tendremos la certeza práctica de afirmar, sin error, cuales datos pertenecen a cada uno ;

3) De las aplicaciones realizadas al capítulo de la transvariación entre mas de dos distribuciones, con la muestra de observaciones correspondientes al año 1956, se desprende la existencia de un moderado valor, alrededor de un medio, para la estimación de los parámetros de la transvariación en la población, en la realización del e.a. considerado en 7.3. Es decir que si disponemos al azar de dos observaciones pertenecientes a una misma rama de industria, correspondiente al año 1956, siendo ellas particulares determinaciones de dos de las cuatro v.a. posibles, no tenemos la certeza práctica para afirmar de qué v.a. proviene cada observación, como tampoco podemos afirmar, con probabilidad aceptable, que dichos valores pueden provenir con la misma o aproximadamente igual probabilidad, de cualquiera de las cuatro v.a. consideradas, que corresponden respectivamente a los casos de probabilidad e intensidad de transvariación próximas a cero o próximas a la unidad ;

4) Resulta fértil en especulaciones la observación comparativa de los parámetros de la transvariación con y sin hipótesis de distribución normal y de ellos entre si. La probabilidad de transvariación resulta invariablemente superior a la intensidad de transvariación de orden uno, salvo cuando ésta es igual a la unidad, tanto en la hipótesis de distribución normal como sin ella, pero la diferencia generalmente es de escasa magnitud cuando trabajamos con la hipótesis de distribución normal, acentuándose

sensiblemente en muchos casos, cuando elaboramos directamente las observaciones sin sujeción a ninguna hipótesis sobre el modelo teórico de distribución estadística. Este comportamiento de los parámetros de la transvariación tiene su explicación. En primer lugar se explica que la probabilidad de transvariación supere a la intensidad de transvariación, porque conforme a la definición de estos parámetros, en aquélla interviene el cincuenta por ciento de la frecuencia a_0 , correspondiente al valor $t = 0$, en cambio para la segunda, al ser nula la medida de la transvariación en el intervalo con centro en el origen cero, es también nulo el sumando del momento de transvariación que contiene como factor dicha medida, generando un distanciamiento tanto mayor entre los dos parámetros, cuanto mayor es el valor relativo de la frecuencia a_0 comparada con el tamaño de la muestra y con la frecuencia a_{-c} y también depende de la magnitud con que la diferencia de las medias supere a la diferencia de los valores medianos. Estas últimas observaciones en torno a las frecuencias a_0 y a_{-c} , acusan su impacto para que los parámetros de la transvariación resulten más altos cuando adoptamos la hipótesis de distribución normal, porque en este supuesto, las frecuencias citadas no intervienen por trabajarse con un modelo teórico de v.a. continua. Damos por admitido que dicho modelo teórico sea representativo del modelo empírico, es decir, que la distribución empírica converja en probabilidad a la distribución normal;

5) Por ultimo, es necesario no perder de vista, en todo el presente trabajo, el sistema de hipótesis que le sirve de base y que es algo así como las reglas de juego de la teoría. Análoga observación corresponde realizar en torno al sistema de hipótesis sobre las que descansan las muestras estudiadas, algunas de las cuales recalamos:

a) que la muestra es representativa de la población, por lo cual no admitimos la existencia de errores en las variables;

b) que el tamaño de la muestra ubica a la misma dentro de la clase de las grandes muestras, razón por la cual trabajamos con las propiedades asintóticas de las distribuciones y no con las distribuciones exactas, de empleo en el dominio de las pequeñas muestras;

c) que nuestro análisis estadístico comprende uno solo de los grandes sectores de la economía: el sector industrial;

para él se deduce, en base a las consideraciones precedentes, el pleno empleo del factor trabajo, situación que es vivida con caracteres agudos por los empresarios industriales, quienes se disputan la mano de obra calificada, de marcada escasez en el mercado del trabajo, situación que facilita notablemente el mejoramiento observado en el nivel de vida de este sector de la población ;

d) el mejoramiento observado en el nivel de vida de los trabajadores de la industria, es también muy relativo ; no significa otra cosa que los obreros industriales, en 1955, 1956 y 1957, años sometidos a estudio, pueden adquirir con sus ingresos, mas bienes que los que podían adquirir en el año base, o sea en 1952 ; pero, en la dinámica social, en un medio que aspire a un inexorable desarrollo económico y social, en armonía con el constante perfeccionamiento de la técnica de producción de bienes de capital, de bienes de consumo y de servicios, que señala el desarrollo de la humanidad y en particular de los pueblos que se han colocado a la vanguardia de la civilización, en este

CONCEPTO	Importe (En millones de m\$n) (año 1957)	%
Agricultura.	23.500	11,0
Ganadería	17.314	8,1
Pesca	250	0,1
Minería	2.243	1,0
Industrias Manufactureras	45.937	21,4
Construcciones	11.547	5,4
Comercio, transporte y comunicaciones . . .	58.736	27,4
Viviendas y finanzas	14.922	7,0
Otros servicios (*)	39.731	18,6
Producto bruto interno según remuneración de los factores	214.180	100,0

(*) Incluye otros servicios públicos, servicios personales y servicios del gobierno.

Años	PRODUCTO BRUTO INTERNO (A precios de mercado del año 1950)		POBLACIÓN AL FIN DE CADA AÑO	
	En millones de m.\$n	Indice Base 1950 = 100	Total	Indice Base 1950 = 100
1902	10.756	17,3	4.871,8	28,0
07	15.890	25,5	5.821,8	33,4
12	19.896	31,9	7.147,4	41,0
17	19.131	30,7	8.445,1	48,5
22	25.491	40,9	9.521,4	54,7
27	33.184	53,3	11.126,4	63,9
32	33.863	54,4	12.517,3	71,8
37	39.754	63,8	13.608,4	78,1
42	45.908	73,7	14.755,7	84,7
45	48.836	78,4	15.520,0	89,1
46	53.197	85,4	15.787,2	90,6
47	59.114	94,9	16.100,4	92,4
48	62.353	100,1	16.519,1	94,8
49	61.544	98,8	16.961,5	97,4
50	62.291	100,0	17.422,2	100,0
51	64.222	103,1	17.855,3	102,5
52	59.986	96,3	18.230,1	104,6
53	63.288	101,6	18.575,5	106,6
54	65.904	105,8	18.942,6	108,7
55	68.894	110,6	19.307,2	110,8
56	68.582	110,1	19.685,7	113,0
57	71.012	114,0	20.061,4	115,1

no solo ya disputado mundo sino disputado sistema planetario de la segunda mitad del siglo XX, para ubicarnos en la realidad de nuestra dimension espacio-tiempo, no puede constituir un límite inferior de nuestra aspiración como Nación, la de mantener un nivel de vida o un poder adquisitivo de los ingresos, pues en la realidad dinámica de la vida, detenerse, es retroceder ;

e) que el mejoramiento observado en la producción industrial no implica el crecimiento de la producción de bienes y servicios por habitante ; el sector industrial integra el conjunto de grandes sectores en la que se divide la economía, a fin de discriminar el ingreso nacional y analizar la contribución de cada sector a la formación de dicho ingreso, en la unidad de tiempo y para un área geográfica determinada. La posición relativa de cada sector de la economía en la formación del producto bruto interno, según remuneración de los factores, para el año 1957, es la siguiente ⁽¹¹⁾.

A fin de apreciar el comportamiento de la producción nacional de bienes y servicios en relación a la población, detallamos a continuación las variaciones en el tiempo operada en la población y en el valor del producto bruto interno a precios constantes, a fin de hacerlos comparables. Dichas series se dan también en números índices, con base 100 en el año 1950 ⁽¹²⁾.

⁽¹¹⁾ Fuente : BANCO CENTRAL DE LA REPÚBLICA ARGENTINA : *Boletín Estadístico* (publicación mensual reiniciada en enero de 1958 por el Departamento de Investigaciones Económicas) - Buenos Aires - 1958.

⁽¹²⁾ Fuente : 1) DIRECCIÓN NACIONAL DE ESTADÍSTICA Y CENSOS : a) *Boletín Mensual de Estadística de Buenos Aires*. b) *Informe Demográfico de la República Argentina 1944-1954* - Buenos Aires. 1956.

2) BANCO CENTRAL DE LA REPÚBLICA ARGENTINA : *Boletín Estadístico* - Buenos Aires - 1958.

3) CEPAL (Comisión Económica para América Latina) : *El desarrollo Económico de la Argentina*. General E/C.N. 12/429 - Santiago de Chile - 1958.

TABLA 1

VARIABLES ESTADÍSTICAS x_i con x_j		a_0	a_{-c}	$\sum_{t=1}^{k-1} a_t$	p_{ij}	$P_{ij} = 2 p_{ij}$
i	j	(1)	(2)	(3)	(4)	(5)
1	2	9	9	26	0,367.470	0,734.940
1	3	2	15	4	0,060.241	0,120.482
1	4	5	7	12	0,174.699	0,349.398
2	3	2	13	3	0,048.193	0,096.386
2	4	7	1	16	0,234.940	0,469.880
3	4	2	5	37	0,457.831	0,915.662

TABLA 2

VARIABLES ESTADÍSTICAS x_i con x_j		$D^2(p_{ij,\Lambda})$	$D^2(p_{ij,M,\Lambda})$	$\mu_{11}(p_{\Lambda}, p_{M,\Lambda})$	$D^2(P_{\Lambda})$	$D(P_{\Lambda})$
i	j	(1)	(2)	(3)	(4)	(5)
1	2	0,002.504	0,002.718	0,002.241	0,002.715	0,0521
1	3	0,000.617	0,002.498	0,000.367	0,002.259	0,0475
1	4	0,001.575	0,002.792	0,001.065	0,004.684	0,0684
2	3	0,000.486	0,002.571	0,000.294	0,001.813	0,0426
2	4	0,001.935	0,003.012	0,001.433	0,005.014	0,0708
3	4	0,002.954	0,002.866	0,002.792	0,000.974	0,0312

TABLA 3

VARIABLES ESTADÍSTICAS x_i con x_j		h	c	$\sum_{t=1}^{k-1} t a_t$	$\sum_{t=1}^{k-1} t^2 a_t$	$\sum_{t=-(c-1)}^{k-1} a_t$	$\sum_{t=-(c-1)}^{k-1} t a_t$	$\sum_{t=-(c-1)}^{k-1} t^2 a_t$
i	j	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	2	3	1	60	212	35	60	212
1	3	3	6	11	43	41	127	473
1	4	3	7	64	562	46	183	1127
2	3	3	5	7	21	42	110	354
2	4	3	6	80	710	44	153	1009
3	4	3	1	267	3050	39	267	3050

TABLA 4

VARIABLES ESTADÍSTICAS x_i con x_j		$m_{ij,1}$	$m_{ij,2}$	$m_{ij,M,1}$	$m_{ij,M,2}$	$I_{ij,1}$
i	j	(1)	(2)	(3)	(4)	(5)
I	2	2,168700	22,988220	3,433775	39,795645	0,631579
I	3	0,397595	4,662705	13,482085	376,594755	0,029491
I	4	2,313280	60,940470	18,253225	644,429205	0,126733
2	3	0,253015	2,277135	11,566400	271,521240	0,021875
2	4	2,891600	76,988850	15,072465	480,258615	0,191847
3	4	9,650715	330,726750	11,060370	392,860005	0,872549

TABLA 5

VARIABLES ESTADÍSTICAS x_i con x_j		$D^2(m_{ij,1,\Lambda})$	$D^2(m_{ij,1,M,\Lambda})$	$\mu_{11}(m_{1,\Lambda}, m_{1,M,\Lambda})$	$D^2(I_{ij,1,\Lambda})$	$D(I_{ij,1,\Lambda})$
j	i	(1)	(2)	(3)	(4)	(5)
I	2	0,222.987	0,341.522	0,268.872	0,001.661	0,0408
I	3	0,054.934	2,375.953	0,078.768	0,000.288	0,0170
I	4	0,677.917	3,795.719	0,820.665	0,001.593	0,0399
2	3	0,026.989	1,679.752	0,038.364	0,000.207	0,0144
2	4	0,836.921	3,086.334	1,042.123	0,002.424	0,0492
3	4	2,897.445	3,299.125	3,084.614	0,000.215	0,0147

TABLA 6

VARIABLES ESTADÍSTICAS x_i con x_j		$D^2(x_i)$	$D^2(x_j)$	r_{ij}	$D^2(x_i - x_j)$	$d = \bar{x}_i - \bar{x}_j $	d_1
i	j	(1)	(2)	(3)	(4)	(5)	(6)
I	2	351,2063	393,7169	0,90054	75,17585	2,6	0,299882
I	3	351,2063	712,9220	0,88158	181,85379	18,3	1,357031
I	4	351,2063	1050,4081	0,54953	734,04352	21,1	0,778792
2	3	393,7169	712,9220	0,90524	147,43363	15,7	1,293008
2	4	393,7169	1050,4081	0,51923	776,19752	18,5	0,664027
3	4	712,9220	1050,4081	0,49650	904,02126	2,8	0,093126

TABLA 7

VARIABLES ESTADÍSTICAS x_i con x_j		$p_{ij} = \Phi(-d_1)$	$P_{ij} = 2\Phi(-d_1)$	$D^2(P_\lambda)$	$D(P_\lambda)$
i	j	(1)	(2)	(3)	(4)
1	2	0,382.134	0,764.268	0,004.394	0,0663
1	3	0,087.388	0,174.776	0,003.518	0,0593
1	4	0,218.052	0,436.104	0,005.998	0,0774
2	3	0,098.007	0,196.014	0,003.843	0,0620
2	4	0,253.339	0,506.678	0,006.096	0,0781
3	4	0,462.902	0,925.804	0,001.675	0,0409

TABLA 8

VARIABLES ESTADÍSTICAS x_i con x_j		$m_{ij;1}$	$m_{ij;2}$	$m_{ij;1,M}$	$m_{ij;2,M}$	$I_{ij;1}$
i	j	(1)	(2)	(3)	(4)	(5)
1	2	2,313.223	22,712.868	3,458.870	37,587.924	0,668.780
1	3	0,543.158	5,952.055	5,379.887	90,926.895	0,100.961
1	4	3,380.283	88,735.690	10,808.684	367,021758	0,312.738
2	3	0,561.045	5,641.120	4,844.068	73,716.814	0,115.821
2	4	4,228.727	118,409.663	11,114.708	388,098.760	0,380.462
3	4	10,646.826	388,662.134	11,995.028	452,010.630	0,887.603

TABLA 9

VARIABLES ESTADÍSTICAS x_i con y_j		$D^2(m_{1,\lambda})$	$D^2(m_{1,M,\lambda})$	$\mu_{11}(m_{1,\lambda}, m_{1,M,\lambda})$	$D^2(I_{1,\lambda})$	$D(I_{1,\lambda})$
i	j	(1)	(2)	(3)	(4)	(5)
1	2	0,211.728	0,312.486	0,252.755	0,001.121	0,033.481
1	3	0,068.987	0,755.891	0,158.166	0,001.546	0,039.319
1	4	0,942.788	3,051.117	1,506.366	0,002.559	0,050.587
2	3	0,064.955	0,612.821	0,143.069	0,001.706	0,041.304
2	4	1,225.933	3,226.334	1,824.860	0,002.464	0,049.639
3	4	3,357.372	3,757.644	3,345.870	0,002.628	0,051.264

TABLA 10

VARIABLES ESTADÍSTICAS	h_i	MEDIAS			MEDIANAS		
x_i con y_i		\bar{x}_i	\bar{y}_i	c_i	M_{x_i}	M_{y_i}	c_i
i	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0,20	4,89	4,83	0	4,90	4,90	0
2	0,20	4,61	4,61	0	4,57	4,63	0
3	0,20	3,68	4,01	2	3,59	3,90	2
4	0,30	2,74	2,97	1	2,55	2,78	1
12	0,10	0,29	— 0,23	5	0,31	— 0,22	5
13	0,10	1,20	— 0,82	20	1,19	— 0,82	20
14	0,30	2,02	— 1,87	13	2,23	— 1,95	14
23	0,30	— 8,28	8,61	56	— 8,10	8,40	55
24	0,40	— 7,32	7,57	37	— 7,05	7,03	35
34	0,40	— 6,43	6,96	33	— 6,15	6,68	32
123	0,30	— 3,40	3,78	24	— 3,26	3,66	23
124	0,30	— 2,43	2,74	17	— 2,25	2,48	16
134	0,30	— 1,52	2,13	12	— 1,44	1,97	11
234	0,50	— 11,01	11,58	45	— 10,64	11,17	44
1234	0,40	— 6,14	6,74	32	— 5,86	6,42	31

TABLA 11

VARIABLES ESTADÍSTICAS	p_i	P_i	$D^2(p_{i,\Lambda})$	$D^2(p_{i,M,\Lambda})$	μ_{11}	$D^2(P_{i,\Lambda})$	$D(P_{i,\Lambda})$
x_i con y_i							
i	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	0,439759	0,879518	0,002674	0,002718	0,002681	0,000239	0,0155
2	0,481928	0,963856	0,002971	0,002975	0,002939	0,000283	0,0168
3	0,259036	0,518072	0,001937	0,002498	0,001579	0,003882	0,0623
4	0,331325	0,662650	0,001930	0,002094	0,002020	0,000689	0,0262
12	0,096386	0,192772	0,000915	0,002755	0,000588	0,003164	0,0562
13	0,012048	0,024096	0,000145	0,002681	0,000074	0,000573	0,0239
14	0,012048	0,024096	0,000145	0,002792	0,000074	0,000573	0,0239
23	0	0	0	0,002939	0	0	0
24	0	0	0	0,002755	0	0	0
34	0	0	0	0,002865	0	0	0
123	0	0	0	0,002902	0	0	0
124	0	0	0	0,003012	0	0	0
134	0,012048	0,024096	0,000145	0,002718	0,000074	0,000573	0,0239
234	0	0	0	0,003012	0	0	0
1234	0	0	0	0,002865	0	0	0

TABLA 12

VARIABLES ESTADÍSTICAS	$m_{i,1}$	$m_{i,2}$	$m_{i,M,1}$	$m_{i,M,2}$	$I_{i,1}$
x_i con y_i					
i	(1)	(2)	(3)	(4)	(5)
I	0,034.704	0,007.558	0,034.704	0,007.558	I
2	0,037.114	0,006.536	0,037.114	0,006.536	I
3	0,019.762	0,003.066	0,224.113	0,103.591	0,088.179
4	0,040.115	0,014.341	0,177.447	0,103.663	0,226.068
I2	0,004.700	0,000.744	0,223.407	0,119.446	0,021.038
I3	0,000.964	0,000.077	0,957.252	2,064.502	0,001.007
I4	0,023.852	0,047.219	2,022.756	9,735.306	0,011.792
23	0	0	10,897.656	229,481.933	0
24	0	0	10,756.972	213,656.718	0
34	0	0	9,573.425	169,232.514	0
I23	0	0	4,640.737	41,606.836	0
I24	0	0	3,235.976	20,232.426	0
I34	0,001.084	0,000.098	2,339.704	10,130.240	0,000.463
234	0	0	16,584.072	531,441.798	0
I234	0	0	9,668.842	165,299.625	0

TABLA 13

VARIABLES ESTADÍSTICAS	$D^2(m_{i,1,\Lambda})$	$D^2(m_{i,M,1,\Lambda})$	μ_{11}	$D^3(I_{i,1,\Lambda})$	$D(I_{i,1,\Lambda})$
x_i con y_i					
i	(1)	(2)	(3)	(4)	(5)
I	0,000.077	0,000.077	0,000.077	0	0
2	0,000.063	0,000.063	0,000.063	0,000.001	0,0010
3	0,000.033	0,000.651	0,000.080	0,000.476	0,0218
4	0,000.155	0,000.880	0,000.235	0,002.977	0,0546
I2	0,000.009	0,000.848	0,000.025	0,000.168	0,0130
I3	0,000.001	0,014.002	0,000.013	0,000.001	0,0010
I4	0,000.569	0,068.826	0,001.122	0,000.135	0,0116
23	0	1,350.281	0	0	0
24	0	1,194.442	0	0	0
34	0	0,946.123	0	0	0
I23	0	0,244.761	0	0	0
I24	0	0,119.035	0	0	0
I34	0,000.001	0,056.781	0,000.018	0,000.002	0,0014
234	0	3,126.956	0	0	0
I234	0	0,875.770	0	0	0

TABLA 14

VARIABLES ESTADÍSTICAS	$D^2(x_i)$	$D^2(y_i)$	r	$D^2(x - y)$	$d = \bar{x} - \bar{y} $	\bar{d}_1
x_i con y_i						
i	(1)	(2)	(3)	(4)	(5)	(6)
1	0,607.900	0,801.100	0,88.593	0,172.391	0,06	0,144.551
2	0,607.900	1,337.900	0,69.222	0,697.210	0	0
3	0,597.600	0,989.900	0,73.046	0,463.776	0,33	0,484.574
4	0,762.400	1,499.100	0,76.410	0,627.701	0,23	0,290.303
12	0,083.217	0,083.817	-0,77.219	0,296.014	0,52	0,955.756
13	0,219.495	0,215.741	-0,43.497	0,624.544	2,02	2,556.053
14	0,816.986	1,358.295	-0,84.544	3,956.502	3,89	1,955.662
23	2,222.531	3,811.947	-0,80.630	10,728.273	16,89	5,156.616
24	2,173.435	3,771.393	-0,77.977	10,409.818	14,89	4,615.015
34	2,055.729	3,666.652	-0,76.121	9,902.171	13,39	4,255.156
123	0,746.967	1,090.480	-0,74.031	3,173.747	7,18	4,030.309
124	0,857.072	1,653.391	-0,74.120	4,275.139	5,17	2,500.434
134	0,858.324	1,705.910	-0,66.260	4,167.796	3,65	1,787.885
234	4,463.714	7,654.416	-0,76.454	21,056.008	22,59	4,922.981
1234	2,160.400	3,802.400	-0,76.066	10,323.150	12,88	4,008.750

TABLA 15

VARIABLES ESTADÍSTICAS	$p_i = \Phi(-d_1)$	$P_i = 2 \Phi(-d_1)$	$D^2(P_i)$	$D(P_i)$
x_i con y_i				
i	(1)	(2)	(3)	(4)
1	0,442.533	0,885.066	0,002.481	0,049.810
2	0,500.000	1,000.000	0	0
3	0,313.991	0,627.982	0,005.698	0,075.485
4	0,385.792	0,771.584	0,004.299	0,065.567
12	0,169.601	0,339.202	0,005.467	0,073.939
13	0,005.293.8	0,010.587.6	0,000.255.5	0,015.984
14	0,025.254	0,050.508	0,001.170	0,034.205
23	0,000.000.126	0,000.000.252	0,000.000.006	0,000.078
24	0,000.001.966	0,000.003.932	0,000.000.096	0,000.310
34	0,000.010.442	0,000.020.884	0,000.000.509	0,000.226
123	0,000.027.863	0,000.055.726	0,000.001.359	0,001.166
124	0,006.202.000	0,012.404	0,000.299.000	0,017.292
134	0,036.899.000	0,073.798	0,001.667.000	0,040.829
234	0,000.000.426	0,000.000.852	0,000.000.021	0,000.144
1234	0,000.030.660	0,000.061.320	0,000.001.500	0,001.225

TABLA 16

VARIABLES ESTADÍSTICAS	m_1	m_2	$m_{1,M}$	$m_{2,M}$	I_1
x_i con y_i					
i	(1)	(2)	(3)	(4)	(5)
1	0,137.365	0,068.047	0,165.642	0,086.196	0,829.288
2	0,333.113	0,348.605	0,333.113	0,348.605	1,000.000
3	0,137.969	0,100.092	0,271.685	0,231.888	0,507.827
4	0,214.297	0,192.873	0,316.073	0,313.851	0,677.998
12	0,049.278	0,024.580	0,217.054	0,148.007	0,227.031
13	0,001.329.4	0,000.620.7	0,315.278	0,312.272	0,004.216.6
14	0,019.003	0,025.997	0,793.538	1,978.251	0,023.947
23	0,000.000.075	0,000.000.100	1,306.703	5,364.137	0,000.000.057
24	0,000.001.264	0,000.001.641	1,287.163	5,204.909	0,000.000.982
34	0,000.007.039	0,000.009.152	1,255.385	4,951.086	0,000.005.607
123	0,000.011.132	0,000.008.505	0,710.719	1,586.874	0,000.015.663
124	0,004.138.000	0,005.123.000	0,824.873	2,137.570	0,005.016.000
134	0,030.041.000	0,044.139.000	0,814.451	2,083.898	0,036.885.000
234	0,000.000.370	0,000.000.613	1,830.628	10,528.004	0,000.000.202
1234	0,000.020.090	0,000.057.760	1,281.794	5,161.575	0,000.015.670

TABLA 17

VARIABLES ESTADÍSTICAS	$D^2(m_{1,\Lambda})$	$D^2(m_{1,M,\Lambda})$	μ_{11}	$D^2(I_{1,\Lambda})$	$D(I_{1,\Lambda})$
x_i con y_i					
i	(1)	(2)	(3)	(4)	(5)
1	0,000.600	0,000.717	0,000.653	0,000.366	0,019.131
2	0,002.898	0,002.898	0,002.898	0	0
3	0,000.988	0,001.928	0,001.319	0,001.972	0,044.407
4	0,001.792	0,002.609	0,002.127	0,001.072	0,032.741
12	0,000.270	0,001.230	0,000.482	0,002.432	0,049.315
13	0,000.007.5	0,002.596	0,000.035.2	0,000.071.6	0,008.462
14	0,000.313	0,016.445	0,001.035	0,000.433	0,020.809
23	0,000.000.001	0,044.593	0,000.000.015	0,000.000.001	0,000.027
24	0,000.000.020	0,043.270	0,000.000.230	0,000.000.012	0,000.110
34	0,000.000.112	0,041.160	0,000.001.153	0,000.000.069	0,000.263
123	0,000.000.104	0,013.192	0,000.000.982	0,000.000.020	0,000.143
124	0,000.062.000	0,017.770	0,000.282.000	0,000.013.000	0,003.605
134	0,000.527.000	0,017.324	0,001.577.000	0,000.655.000	0,025.593
234	0,000.000.007	0,087.644	0,000.000.101	0,000.000.002	0,000.047
1234	0,000.000.700	0,042.909	0,000.003.550	0,000.000.430	0,000.656

TABLA 18

VARIABLES ESTADÍSTICAS x_i con $x_{i+1,\dots,4}$		h	MEDIAS			MEDIANAS		
i	$i+1,\dots,4$		\bar{x}_i	$\bar{x}_{i+1,\dots,4}$	c	M_{x_i}	$M_{x_{i+1,\dots,4}}$	c
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
3	4	3	118,2	121,0	1	113,0	118,5	2
2	34	3	102,5	119,6	6	100,0	115,6	5
1	234	3	99,9	113,9	5	99,7	109,8	3

TABLA 19

VARIABLES ESTADÍSTICAS x_i con $x_{i+1,\dots,4}$		$p_{i(i+1,\dots,4)}$	$P_{i(i+1,\dots,4)}$	$D^2(p_\Lambda)$	$D(p_{M,\Lambda})$	μ_{11}	$D^2(P_\Lambda)$
i	$i+1,\dots,4$	(1)	(2)	(3)	(4)	(5)	(6)
3	4	0,457831	0,915662	0,002954	0,002866	0,002792	0,000974
2	34	0,141557	0,283114	0,000654	0,001332	0,000429	0,002073
1	234	0,196787	0,393574	0,000573	0,000907	0,000397	0,001603

TABLA 20

VARIABLES ESTADÍSTICA x_i con $x_{i+1,\dots,4}$		$m_{i(i+1,\dots,4)1}$	$m_{i(i+1,\dots,4)2}$	$m_{i(i+1,\dots,4)M,1}$	$m_{i(i+1,\dots,4)M,2}$
i	$i+1,\dots,4$	(1)	(2)	(3)	(4)
3	4	9,650,715	330,726.750	11,060.370	392,860.005
2	34	1,572.264	39,631.896	16,463.592	491,576.472
1	234	1,626.480	29,529.648	13,168.464	318,031.056

TABLA 21

VARIABLES ESTADÍSTICAS		$D^2 (m_{1,\Lambda})$	$D^2 (m_{M,1,\Lambda})$	μ_{11}
x_i con $x_{i+1}, \dots, 4$				
i	$i+1, \dots, 4$	(1)	(2)	(3)
3	4	2,897.445	3,299.125	3,084.614
2	34	0,225.211	1,336.525	0,254.834
1	234	0,108.404	0,583.156	0,131.083

TABLA 22

VARIABLES ESTADÍSTICAS		$D^2 (x_i)$	$D^2 (x_{i+1}, \dots, 4)$	r	$D^2 (x_i -$ $- x_{i+1}, \dots, 4)$	$d = \bar{x}_i -$ $- \bar{x}_{i+1}, \dots, 4 $	d_1
x_i con $x_{i+1}, \dots, 4$							
i	$i+1, \dots, 4$	(1)	(2)	(3)	(4)	(5)	(6)
3	4	712,9220	1050,4081	0,4965	904,0213	2,8	0,09313
2	34	393,7169	883,5120	0,6896	463,7870	17,1	0,79403
1	234	351,2063	784,9530	0,7289	370,7377	14,0	0,72710

TABLA 23

VARIABLES ESTADÍSTICAS		$p_{i(i+1, \dots, 4)} =$ $= \Phi(-d_1)$	$D^2 (p_{i(i+1, \dots, 4)\Lambda})$	$D^2 (p_{i(i+1, \dots, 4)M, \Lambda})$	μ_{11}
x_i con $x_{i+1}, \dots, 4$					
i	$i+1, \dots, 4$	(1)	(2)	(3)	(4)
3	4	0,462.902	0,003.032	0,003.049	0,002.823
2	34	0,213.592	0,001.018	0,001.515	0,000.647
1	234	0,233.584	0,000.722	0,001.008	0,000.471

TABLA 24

VARIABLES ESTADÍSTICAS		$m_{i(i+1,\dots,4)1}$	$m_{i(i+1,\dots,4)2}$	$m_{i(i+1,\dots,4)M,1}$	$m_{i(i+1,\dots,4)M,2}$
x_i con $x_{i+1,\dots,4}$					
i	$i+1,\dots,4$	(1)	(2)	(3)	(4)
3	4	10,646.826	388,662.134	11,995.028	452,010.630
2	34	2,616.021	54,327.242	8,591.544	231,893.510
1	234	2,626.953	49,821.047	7,681.487	185,368.838

TABLA 25

VARIABLES ESTADÍSTICAS		$D^2 (m_{1,\Lambda})$	$D^2 (m_{M,1,\Lambda})$	μ_{11}
x con $x_{i+1,\dots,4}$				
i	$i+1,\dots,4$	(1)	(2)	(3)
3	4	3,357.372	3,757.644	3,345.870
2	34	0,287.799	0,958.116	0,464.185
1	234	0,173.054	0,509.498	0,267.820

BIBLIOGRAFIA SOBRE LA TRANSVARIACION

- 1) BOLDRINI, M. — *Su alcune differenze sessuali secondarie nelle dimensioni del corpo umano alla nascita e alle età superiori* - « Archivio per l'Antropologia e l'Etnologia », vol. XLIX, 1919.
- 2) BOLDRINI, M. — *I cadaveri degli sconosciuti. Ricerche demografiche e antropologiche sul materiale della Morgue di Roma* - « Scuola Positiva », 1920.
- 3) BOLDRINI, M. — *Differenze sessuali nei pesi del corpo e degli organi umani* - « Rendiconti della R. Accademia dei Lincei », vol. XXIX, s. V., fasc. 1-4, 1920.
- 4) BOLDRINI, M. — *Gli studi statistici sul sesso : la proporzione dei sessi nelle nascite e i caratteri sessuali secondari* - « Rassegna di studi sessuali », 1921.
- 5) BOLDRINI, M. — *Misure interne ed esterne di alcune ossa lunghe nell'uomo e nella donna* - « Rendiconti della R. Accademia dei Lincei », vol. XXXIII, s.V., fasc. 7 10, 1924.

- 6) BOLDRINI, M. — *Biometria. Problemi della vita della specie e degli individui* - Ed. Cedam, Padova, 1927.
- 7) CASTELLANO, V. — *Sullo scarto quadratico medio della probabilità di transvariazione* - «Metron», vol. XI, n. 4, 1934.
- 8) DAGUM, C. — *Transvariación entre más de dos distribuciones* (En imprenta, a publicarse por la Universidad de Roma en el libro en Homenaje al Prof. GINI); *Transvariazione fra più di due distribuzioni* - (En imprenta, a publicarse en Roma en el Libro Memorie sulla Transvariazione).
- 9) DE LUCIA, L. — *Transvariazione tra caratteri connessi* - «Atti della XII Riunione della Società Italiana di Statistica», 1952.
- 10) FILIPPONI, A. — *Sul grado di stabilità nei caratteri di Protomagalhaensia Marottai* (Sporozoa, Gregarinidae) - «Rivista di parassitologia», vol. XIV, n. 3, 1953.
- 11) GINI, C. — *Il concetto di transvariazione e le sue prime applicazioni* - «Giornale degli Economisti» e «Rivista di Statistica» 1916 y en «Memorie di Metodologia Statistica», vol. I, Milano, 1939.
- 12) GINI, C. — *Per la determinazione della probabilità di transvariazione tra più gruppi* - «Atti della V Riunione della Società».
- 13) GINI, C. — *Della misura sintetica della Transvariazione rispetto ad n caratteri* (transvariazione n-dimensionale) - «Atti della XI Riunione della Società Italiana di Statistica», 1951.
- 14) GINI, C. y LIVADA, G. — *Transvariazione a più dimensioni* - «Atti della VI Riunione della Società Italiana di Statistica», 1943.
- 15) GINI, C. y LIVADA, G. — *Nuovi contributi alla teoria della transvariazione* - «Atti della VII Riunione della Società Italiana di Statistica», 1943.
- 16) GINI, C. y SONNINO, G. — *Contributo alla teoria della transvariazione fra seriazioni correlate* - «Atti della XI Riunione della Società Italiana di Statistica», 1951.
- 17) GINI, C., VITERBO, C., BENEDETTI, C. y HERZEL, A. — *Problemi di transvariazione inversa* - «Metron» Vol. XIX n. 1-2, 1958.
- 18) HERZEL, A. — *Influenza del Raggruppamento in classi sulla probabilità e sull'intensità di transvariazione* - «Metron» Vol. XIX n. 1-2, 1958.
- 19) LIVADA, G. — *Procedimento per il calcolo dell'intensità di transvariazione* - «Atti della VI Riunione della Società Italiana di Statistica», 1943.
- 20) MIANI CALABRESE, D. — *La transvariazione rispetto al sesso dei caratteri fisici dell'infanzia* - «Statistica» Ottobre-dicembre 1943.
- 21) OTTAVIANI, G. — *Sulla probabilità che una prova su due variabili casuali X e Y verifichi la disuguaglianza $X < Y$ e sul corrispondente scarto quadratico medio* - Giornale dell'Istituto Italiano degli Attuari - Año X - N° 3-4 - 1939.



CUADRO N. 10

[illegible]

x_3	y_3	FRECUENCIAS																								$(x - y)$	Frecuencias Diag.	$(x-y) \times$ Frecuencias Diag.	$(x-y)^2 \times$ Frecuencias Diag.				
		De 1,90 a menos de 2,10	de 2,10 a menos de 2,30	de 2,30 a menos de 2,50	de 2,50 a menos de 2,70	de 2,70 a menos de 2,90	de 2,90 a menos de 3,10	de 3,10 a menos de 3,30	de 3,30 a menos de 3,50	de 3,50 a menos de 3,70	de 3,70 a menos de 3,90	de 3,90 a menos de 4,10	de 4,10 a menos de 4,30	de 4,30 a menos de 4,50	de 4,50 a menos de 4,70	de 4,70 a menos de 4,90	de 4,90 a menos de 5,10	de 5,10 a menos de 5,30	de 5,30 a menos de 5,50	de 5,50 a menos de 5,70	de 5,70 a menos de 5,90	de 5,90 a menos de 6,10	de 6,10 a menos de 6,30	de 6,30 a menos de 6,50	de 6,50 a menos de 6,70								
		2,00	2,20	2,40	2,60	2,80	3,00	3,20	3,40	3,60	3,80	4,00	4,20	4,40	4,60	4,80	5,00	5,20	5,40	5,60	5,80	6,00	6,20	6,40	6,60								
De 0,70 a menos de 0,90	0,80											1															1	—	5,80	—	—		
de 0,90 a menos de 1,10	1,00																										—	—	5,60	—	—		
de 1,10 a menos de 1,30	1,20																										—	—	—	5,40	—	—	
de 1,30 a menos de 1,50	1,40																										—	—	—	—	5,20	—	
de 1,50 a menos de 1,70	1,60																										—	—	—	—	5,00	—	
de 1,70 a menos de 1,90	1,80																										—	—	—	—	4,80	—	
de 1,90 a menos de 2,10	2,00	1																									1	—	—	—	—	4,60	
de 2,10 a menos de 2,30	2,20																										—	—	—	—	—	4,40	
de 2,30 a menos de 2,50	2,40																										—	—	—	—	—	4,20	
de 2,50 a menos de 2,70	2,60				1	2																					3	—	—	—	—	4,00	
de 2,70 a menos de 2,90	2,80			1																							1	—	—	—	—	3,80	
de 2,90 a menos de 3,10	3,00			1			2	1	2																		6	—	—	—	—	3,60	
de 3,10 a menos de 3,30	3,20					1	1	4	2	2	1																11	—	—	—	—	3,40	
de 3,30 a menos de 3,50	3,40					1	1	3	1	3	3																12	—	—	—	—	3,20	
de 3,50 a menos de 3,70	3,60	1					1			1	3	1	2	2	2						2						15	—	—	—	—	3,00	
de 3,70 a menos de 3,90	3,80									1		1	3	3	1	1											10	—	—	—	—	2,80	
de 3,90 a menos de 4,10	4,00													2	3												5	—	—	—	—	2,60	
de 4,10 a menos de 4,30	4,20												1			2											3	—	—	—	—	2,40	
de 4,30 a menos de 4,50	4,40																		1							1	—	—	—	—	2,20		
de 4,50 a menos de 4,70	4,60									1				1		1		1	1								5	—	2	—	4,40	9,68	
de 4,70 a menos de 4,90	4,80																	2							1	1	4	—	—	—	—	2,00	
de 4,90 a menos de 5,10	5,00													1													1	—	1,80	1	—	1,80	3,24
de 5,10 a menos de 5,30	5,20																					1					1	—	1,60	1	—	1,60	2,56
de 5,30 a menos de 5,50	5,40																			1							1	—	1,40	—	—	—	
de 5,50 a menos de 5,70	5,60															1											1	—	1,20	—	—	—	
de 5,70 a menos de 5,90	5,80																							1			1	—	1,00	4	—	4,00	4,00
FRECUENCIAS		2	—	2	1	4	5	8	5	8	7	3	6	9	6	5	—	3	2	1	2	1	—	2	1	83							
$(x - y)$		3,80	3,60	3,40	3,20	3,00	2,80	2,60	2,40	2,20	2,00	1,80	1,60	1,40	1,20	1,00	0,80	0,60	0,40	0,20	0	—	0,20	0,40	0,60	0,80							
Frecuencias Diag.		—	—	—	—	—	—	—	—	—	—	—	1	—	—	1	1	4	3	6	11	14	15	13	5								
$(x-y) \times$ Frecuencias Diag.		—	—	—	—	—	—	—	—	—	—	—	1,60	—	—	1,00	0,80	2,40	1,20	1,20	—	—	2,80	6,00	7,80	—	—	—	—	—	—	—	
$(x-y)^2 \times$ Frecuencias Diag.		—	—	—	—	—	—	—	—	—	—	—	2,56	—	—	1,00	0,64	1,44	0,48	0,24	—	—	0,56	2,40	4,68	3,20							

CUADRO N. II

$x_4 \backslash y_4$		Frecuencias																												$(x - y)$			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias																												Frecuencias Diag.			
		Frecuencias</																															

FRECUENCIAS	(x - y)	Frecuencias Diag.	(x-y) × Frecuencias Diag.	(x-y) ² × Frecuencias Diag.
1	-9,60	—	—	—
1	-9,30	—	—	—
2	-9,00	—	—	—
1	-8,70	—	—	—
12	-8,40	—	—	—
25	-8,10	—	—	—
15	-7,80	—	—	—
10	-7,50	—	—	—
7	-7,20	—	—	—
3	-6,90	—	—	—
1	-6,60	—	—	—
2	-6,30	—	—	—
—	-6,00	—	—	—
1	-5,70	—	—	—
—	-5,40	—	—	—
—	-5,10	—	—	—
—	-4,80	—	—	—
—	-4,50	—	—	—
—	-4,20	—	—	—
1	-3,90	1	-3,90	15,21
1				

$y_{12} = -y_1 + y_2$ $x_{12} = x_1 - x_2$	$y_{12} = -y_1 + y_2$ $x_{12} = x_1 - x_2$																								Frecuencias	$x - y$	Frecuencias Diag.	(x-y) Frecuencias Diag.	(x-y) ² Frecuencias Diag.	
	De -1,15 a menos de -1,05	de -1,05 a menos de -0,95	de -0,95 a menos de -0,85	de -0,85 a menos de -0,75	de -0,75 a menos de -0,65	de -0,65 a menos de -0,55	de -0,55 a menos de -0,45	de -0,45 a menos de -0,35	de -0,35 a menos de -0,25	de -0,25 a menos de -0,15	de -0,15 a menos de -0,05	de -0,05 a menos de 0,05	de 0,05 a menos de 0,15	de 0,15 a menos de 0,25	de 0,25 a menos de 0,35	de 0,35 a menos de 0,45	de 0,45 a menos de 0,55	de 0,55 a menos de 0,65	de 0,65 a menos de 0,75	de 0,75 a menos de 0,85	de 0,85 a menos de 0,95	de 0,95 a menos de 1,05	de 1,05 a menos de 1,15							
	-1,10	-1,00	-0,90	-0,80	-0,70	-0,60	-0,50	-0,40	-0,30	-0,20	-0,10	0	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00	1,10							
De -1,15 a menos de -1,05	-1,10																						1	1	-2,20	1	-2,20	4,84		
de -1,05 a menos de -0,95	-1,00																							—	-2,10	—	—	—	—	
de -0,95 a menos de -0,85	-0,90																							—	-2,00	—	—	—	—	
de -0,85 a menos de -0,75	-0,80																							—	-1,90	—	—	—	—	
de -0,75 a menos de -0,65	-0,70																							—	-1,80	—	—	—	—	
de -0,65 a menos de -0,55	-0,60																							—	-1,70	—	—	—	—	
de -0,55 a menos de -0,45	-0,50																	1						1	-1,60	—	—	—	—	
de -0,45 a menos de -0,35	-0,40																							—	-1,50	—	—	—	—	
de -0,35 a menos de -0,25	-0,30									1														1	-1,40	—	—	—	—	
de -0,25 a menos de -0,15	-0,20																							—	-1,30	—	—	—	—	
de -0,15 a menos de -0,05	-0,10													1										1	-1,20	—	—	—	—	
de -0,05 a menos de 0,05	0										2		4	1	1									8	-1,10	1	-1,10	1,21		
de 0,05 a menos de 0,15	0,10									2	1	1	5											9	-1,00	—	—	—	—	
de 0,15 a menos de 0,25	0,20								1	1	3	3	4											12	-0,90	—	—	—	—	
de 0,25 a menos de 0,35	0,30						1	2	2	3	5	2												15	-0,80	—	—	—	—	
de 0,35 a menos de 0,45	0,40					1			3	3	3	2				1								13	-0,70	—	—	—	—	
de 0,45 a menos de 0,55	0,50					2			1	1	3	3				1								11	-0,60	—	—	—	—	
de 0,55 a menos de 0,65	0,60								2	1														3	-0,50	—	—	—	—	
de 0,65 a menos de 0,75	0,70			1		1			2															4	-0,40	—	—	—	—	
de 0,75 a menos de 0,85	0,80				1					1														3	-0,30	—	—	—	—	
de 0,85 a menos de 0,95	0,90				1																			1	-0,20	2	-0,40	0,08		
de 0,95 a menos de 1,05	1,00																							—	-0,10	2	-0,20	0,02		
de 1,05 a menos de 1,15	1,10																							—						
Frecuencias	—	—	1	2	4	1	7	9	12	18	9	13	3	2	—	—	—	1	—	—	—	—	1	83						
$x - y$	2,20	2,10	2,00	1,90	1,80	1,70	1,60	1,50	1,40	1,30	1,20	1,10	1,00	0,90	0,80	0,70	0,60	0,50	0,40	0,30	0,20	0,10	0							
Frecuencias Diag.	—	—	—	—	—	1	2	—	1	—	5	3	2	3	8	8	7	8	8	4	8	5	4							
(x-y) Frecuencias Diag.	—	—	—	—	—	1,70	3,20	—	1,40	—	6,00	3,30	2,00	2,70	6,40	5,60	4,20	4,00	3,20	1,20	1,60	0,50	—							
(x-y) ² Frecuencias Diag.	—	—	—	—	—	2,89	5,12	—	1,96	—	7,20	3,63	2,00	2,43	5,12	3,92	2,52	2,00	1,28	0,36	0,32	0,05	—							

CUADRO N. 13

[illegible]

$y_{23}=y_2+y_3$ $x_{23}=x_2+x_3$		FRECUENCIAS																				FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS		FRECUENCIAS</	
--------------------------------------	--	-------------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	-------------	--	---------------	--

$y_{24} = y_2 + y_4$ $x_{24} = x_2 + x_4$		De 3,00 a menos de 3,40	de 3,40 a menos de 3,80	de 3,80 a menos de 4,20	de 4,20 a menos de 4,60	de 4,60 a menos de 5,00	de 5,00 a menos de 5,40	de 5,40 a menos de 5,80	de 5,80 a menos de 6,20	de 6,20 a menos de 6,60	de 6,60 a menos de 7,00	de 7,00 a menos de 7,40	de 7,40 a menos de 7,80	de 7,80 a menos de 8,20	de 8,20 a menos de 8,60	de 8,60 a menos de 9,00	de 9,00 a menos de 9,40	de 9,40 a menos de 9,80	de 9,80 a menos de 10,20	de 10,20 a menos de 10,60	de 10,60 a menos de 11,00	de 11,00 a menos de 11,40	de 11,40 a menos de 11,80	de 11,80 a menos de 12,20	de 12,20 a menos de 12,60	de 12,60 a menos de 13,00	de 13,00 a menos de 13,40	de 13,40 a menos de 13,80	de 13,80 a menos de 14,20	de 14,20 a menos de 14,60	de 14,60 a menos de 15,00	de 15,00 a menos de 15,40	de 15,40 a menos de 15,80	FRECUENCIAS	$(x - y)$	Frecuencias Diag.	$(x-y) \times$ Frecuencias Diag.	$(x-y)^2 \times$ Frecuencias Diag.	
		3,20	3,60	4,00	4,40	4,80	5,20	5,60	6,00	6,40	6,80	7,20	7,60	8,00	8,40	8,80	9,20	9,60	10,00	10,40	10,80	11,20	11,60	12,00	12,40	12,80	13,20	13,60	14,00	14,40	14,80	15,20	15,60						
De -13,80 a menos de -13,40	-13,60																	1																1	-29,20	—	—	—	
de -13,40 a menos de -13,00	-13,20																																	—	-28,80	—	—	—	
de -13,00 a menos de -12,60	-12,80																																	—	-28,40	—	—	—	
de -12,60 a menos de -12,20	-12,40																																	—	-28,00	—	—	—	
de -12,20 a menos de -11,80	-12,00																																	—	-27,60	—	—	—	
de -11,80 a menos de -11,40	-11,60																																	—	-27,20	—	—	—	
de -11,40 a menos de -11,00	-11,20																		1														1	2	-26,80	1	-26,80	718,24	
de -11,00 a menos de -10,60	-10,80																																	—	-26,40	—	—	—	
de -10,60 a menos de -10,20	-10,40																									1							1	-26,00	—	—	—		
de -10,20 a menos de -9,80	-10,00																			1													—	-25,60	—	—	—		
de -9,80 a menos de -9,40	-9,60																																	—	-25,20	—	—	—	
de -9,40 a menos de -9,00	-9,20																1	2	1		1												5	-24,80	—	—	—		
de -9,00 a menos de -8,60	-8,80																																	1	-24,40	—	—	—	
de -8,60 a menos de -8,20	-8,40													1	1	2	1	1															6	-24,00	—	—	—		
de -8,20 a menos de -7,80	-8,00								1				1		2					1													5	-23,60	—	—	—		
de -7,80 a menos de -7,40	-7,60											1	2	3	3	1	1								1								12	-23,20	2	-46,40	1076,48		
de -7,40 a menos de -7,00	-7,20									1	2	1	3			1				1													9	-22,80	—	—	—		
de -7,00 a menos de -6,60	-6,80						2	2	1	5	4	1																					15	-22,40	—	—	—		
de -6,60 a menos de -6,20	-6,40	1				1			3	6		1	1																				13	-22,00	—	—	—		
de -6,20 a menos de -5,80	-6,00					1			1	1																							3	-21,60	—	—	—		
de -5,80 a menos de -5,40	-5,60				1		1	1	1	1		1																					6	-21,20	1	-21,20	449,44		
de -5,40 a menos de -5,00	-5,20																																	—	-20,80	—	—	—	
de -5,00 a menos de -4,60	-4,80									1																								1	-20,40	1	-20,40	416,16	
de -4,60 a menos de -4,20	-4,40							1																										1	-20,00	1	-20,00	400,00	
de -4,20 a menos de -3,80	-4,00		1																															1					
FRECUENCIAS		1	1	—	1	—	3	4	9	10	7	8	8	4	6	5	3	4	2	3	1	—	—	1	—	1	—	—	—	—	—	—	—	1	83				
$(x - y)$		-7,20	-7,60	-8,00	-8,40	-8,80	-9,20	-9,60	-10,00	-10,40	-10,80	-11,20	-11,60	-12,00	-12,40	-12,80	-13,20	-13,60	-14,00	-14,40	-14,80	-15,20	-15,60	-16,00	-16,40	-16,80	-17,20	-17,60	-18,00	-18,40	-18,80	-19,20	-19,60						
Frecuencias Diag.		—	1	—	—	—	—	1	2	—	2	2	2	2	6	9	1	7	8	2	4	2	4	4	4	2	2	3	1	2	2	1	1						
$(x-y) \times$ Frecuencias Diag.		—	-7,60	—	—	—	—	-9,60	-20,00	—	-21,60	-22,40	-23,20	-24,00	-74,40	-115,20	-13,20	-95,20	-112,00	-28,80	-59,20	-30,40	-62,40	-64,00	-65,60	-33,60	-34,40	-52,80	-18,00	-36,80	-37,60	-19,20	-19,60						
$(x-y)^2 \times$ Frecuencias Diag.		—	57,76	—	—	—	—	92,16	200,00	—	233,28	250,88	269,12	288,00	922,56	1474,56	174,24	1294,72	1568,00	414,72	876,16	462,08	973,44	1024,00	1075,84	564,48	591,68	929,28	324,00	677,12	706,88	368,64	384,16						

$y_{34}=y_3+y_4$		$x_{34}=x_3+x_4$																																									
		De 3,00 a menos de 3,40	de 3,40 a menos de 3,80	de 3,80 a menos de 4,20	de 4,20 a menos de 4,60	de 4,60 a menos de 5,00	de 5,00 a menos de 5,40	de 5,40 a menos de 5,80	de 5,80 a menos de 6,20	de 6,20 a menos de 6,60	de 6,60 a menos de 7,00	de 7,00 a menos de 7,40	de 7,40 a menos de 7,80	de 7,80 a menos de 8,20	de 8,20 a menos de 8,60	de 8,60 a menos de 9,00	de 9,00 a menos de 9,40	de 9,40 a menos de 9,80	de 9,80 a menos de 10,20	de 10,20 a menos de 10,60	de 10,60 a menos de 11,00	de 11,00 a menos de 11,40	de 11,40 a menos de 11,80	de 11,80 a menos de 12,20	de 12,20 a menos de 12,60	de 12,60 a menos de 13,00	de 13,00 a menos de 13,40	de 13,40 a menos de 13,80	de 13,80 a menos de 14,20	de 14,20 a menos de 14,60	de 14,60 a menos de 15,00	de 15,00 a menos de 15,40	Frecuencias		$x - y$		Frecuencias Diag.		(x-y) Frecuencias Diag.		(x-y) ² Frecuencias Diag.		
		3,20	3,60	4,00	4,40	4,80	5,20	5,60	6,00	6,40	6,80	7,20	7,60	8,00	8,40	8,80	9,20	9,60	10,00	10,40	10,80	11,20	11,60	12,00	12,40	12,80	13,20	13,60	14,00	14,40	14,80	15,20											
De -12,60 a menos de -12,20	-12,40															1																		1									
de -12,20 a menos de -11,80	-12,00																																										
de -11,80 a menos de -11,40	-11,60																																										
de -11,40 a menos de -11,00	-11,20																																										
de -11,00 a menos de -10,60	-10,80																																										
de -10,60 a menos de -10,20	-10,40																																										
de -10,20 a menos de -9,80	-10,00															1																											
de -9,80 a menos de -9,40	-9,60																																										
de -9,40 a menos de -9,00	-9,20																	1							1																		
de -9,00 a menos de -8,60	-8,80																																										
de -8,60 a menos de -8,20	-8,40															1	2																										
de -8,20 a menos de -7,80	-8,00														1	1		1																									
de -7,80 a menos de -7,40	-7,60															1		1																									
de -7,40 a menos de -7,00	-7,20								1			2	1	2	2	1																											
de -7,00 a menos de -6,60	-6,80							1			1		2	2	1			1																									
de -6,60 a menos de -6,20	-6,40							1		2		3	2																														
de -6,20 a menos de -5,80	-6,00	1					1	3		5	3	2																															
de -5,80 a menos de -5,40	-5,60					2	3	3	5	1	1																																
de -5,40 a menos de -5,00	-5,20			1			2	2																																			
de -5,00 a menos de -4,60	-4,80				1	1	1				1																																
de -4,60 a menos de -4,20	-4,40																																										
de -4,20 a menos de -3,80	-4,00					1	1																																				
de -3,80 a menos de -3,40	-3,60	1																																									
de -3,40 a menos de -3,00	-3,20											1																															
Frecuencias		2	—	1	1	4	8	10	6	9	5	8	5	4	4	6	2	4	—	2	—	—	—	1	—	—	—	—	—	—	—	—	—	1	83								
$x - y$		-6,40	-6,80	-7,20	-7,60	-8,00	-8,40	-8,80	-9,20	-9,60	-10,00	-10,40	-10,80	-11,20	-11,60	-12,00	-12,40	-12,80	-13,20	-13,60	-14,00	-14,40	-14,80	-15,20	-15,60	-16,00	-16,40	-16,80	-17,20	-17,60	-18,00	-18,40											
Frecuencias Diag.		—	1	—	—	—	—	1	4	1	1	5	5	5	8	2	7	5	3	4	2	4	3	3	2	1	3	2	2	3	1	—											
(x-y) Frecuencias Diag.		—	-6,80	—	—	—	—	-8,80	-36,80	9,60	-10,00	-52,00	-54,00	-56,00	-92,80	-24,00	-86,80	-64,00	-39,60	-54,40	-28,00	-57,60	-44,40	-45,60	-31,20	-16,00	-49,20	-33,60	-34,40	-52,80	-18,00	—											
(x-y) ² Frecuencias Diag.		—	46,24	—	—	—	—	77,44	338,56	92,16	100,00	540,80	583,20	627,20	1076,48	288,00	1076,32	819,20	522,72	739,84	392,00	829,44	657,12	693,12	486,72	256,00	806,88	564,48	591,68	929,28	324,00	—											

CUADRO N. 19

$y_{123} = -y_1 + y_2 + y_3$ $x_{123} = -x_1 - x_2 - x_3$		$x - y$																				Frecuencias	$x - y$	Frecuencias Diag.	$(x-y) \times$ Frecuencias Diag.	$(x-y)^2 \times$ Frecuencias Diag.			
		De 1,05 a menos de 1,35	de 1,35 a menos de 1,65	de 1,65 a menos de 1,95	de 1,95 a menos de 2,25	de 2,25 a menos de 2,55	de 2,55 a menos de 2,85	de 2,85 a menos de 3,15	de 3,15 a menos de 3,45	de 3,45 a menos de 3,75	de 3,75 a menos de 4,05	de 4,05 a menos de 4,35	de 4,35 a menos de 4,65	de 4,65 a menos de 4,95	de 4,95 a menos de 5,25	de 5,25 a menos de 5,55	de 5,55 a menos de 5,85	de 5,85 a menos de 6,15	de 6,15 a menos de 6,45	de 6,45 a menos de 6,75	de 6,75 a menos de 7,05						de 7,05 a menos de 7,35	de 7,35 a menos de 7,65	
		1,20	1,50	1,80	2,10	2,40	2,70	3,00	3,30	3,60	3,90	4,20	4,50	4,80	5,10	5,40	5,70	6,00	6,30	6,60	6,90	7,20	7,50						
De -6,15 a menos de -5,85	- 6,00																						1	1	-13,50	1	-13,50	182,25	
de -5,85 a menos de -5,55	- 5,70														1									1	-13,20	—	—	—	
de -5,55 a menos de -5,25	- 5,40																1							1	-12,90	—	—	—	
de -5,25 a menos de -4,95	- 5,10																		1					1	-12,60	—	—	—	
de -4,95 a menos de -4,65	- 4,80											1			1		1							3	-12,30	—	—	—	
de -4,65 a menos de -4,35	- 4,50									1			1	2		1								5	-12,00	—	—	—	
de -4,35 a menos de -4,05	- 4,20													4										4	-11,70	—	—	—	
de -4,05 a menos de -3,75	- 3,90											1	2	1						1				5	-11,40	1	-11,40	129,96	
de -3,75 a menos de -3,45	- 3,60							1	1	2	2	3	1											10	-11,10	1	-11,10	123,21	
de -3,45 a menos de -3,15	- 3,30					1	1	1	4	2	3	4	1											17	-10,80	1	-10,80	116,64	
de -3,15 a menos de -2,85	- 3,00	1				1	2	5	2	5	1	2												19	-10,50	1	-10,50	110,25	
de -2,85 a menos de -2,55	- 2,70				1		2	3	2	1				1										10	-10,20	1	-10,20	104,04	
de -2,55 a menos de -2,25	- 2,40				1		1	1																3	- 9,90	2	-19,80	196,02	
de -2,25 a menos de -1,95	- 2,10						1																	1	- 9,60	—	—	—	
de -1,95 a menos de -1,65	- 1,80			1																				1	- 9,30	2	-18,60	172,98	
de -1,65 a menos de -1,35	- 1,50																							—	- 9,00	6	-54,00	486,00	
de -1,35 a menos de -1,05	- 1,20																							—	- 8,70	1	- 8,70	75,69	
de -1,05 a menos de -0,75	- 0,90																							—	- 8,40	2	-16,80	141,12	
de -0,75 a menos de -0,45	- 0,60																							—	- 8,10	3	-24,30	196,83	
de -0,45 a menos de -0,15	- 0,30																							—	- 7,80	4	-31,20	243,36	
de -0,15 a menos de 0,15	0								1															1					
Frecuencias		1	—	1	2	2	7	11	10	11	6	11	5	8	2	1	2	—	2	—	—	—	—	1	83				
$x - y$		- 1,20	- 1,50	- 1,80	- 2,10	- 2,40	- 2,70	- 3,00	- 3,30	- 3,60	- 3,90	- 4,20	- 4,50	- 4,80	- 5,10	- 5,40	- 5,70	- 6,00	- 6,30	- 6,60	- 6,90	- 7,20	- 7,50						
Frecuencias Diag.		—	—	—	—	—	—	—	1	1	—	1	1	2	1	4	6	8	4	10	4	7	7						
$(x-y) \times$ Frecuencias Diag.		—	—	—	—	—	—	—	- 3,30	- 3,60	—	- 4,20	- 4,50	- 9,60	- 5,10	-20,80	-34,20	-48,00	-25,20	-66,00	-27,60	-50,40	-52,50						
$(x-y)^2 \times$ Frecuencias Diag.		—	—	—	—	—	—	—	10,89	12,96	—	17,64	20,25	46,08	26,01	112,32	194,94	288,00	158,76	435,60	190,44	362,88	393,75						

$y_{124} = -y_1 + y_2 + y_4$	$x_{124} = x_1 - x_2 - x_4$	De 0,15 a menos de 0,45																				de 0,45 a menos de 0,75																				de 0,75 a menos de 1,05																				de 1,05 a menos de 1,35																				de 1,35 a menos de 1,65																				de 1,65 a menos de 1,95																				de 1,95 a menos de 2,25																				de 2,25 a menos de 2,55																				de 2,55 a menos de 2,85																				de 2,85 a menos de 3,15																				de 3,15 a menos de 3,45																				de 3,45 a menos de 3,75																				de 3,75 a menos de 4,05																				de 4,05 a menos de 4,35																				de 4,35 a menos de 4,65																				de 4,65 a menos de 4,95																				de 4,95 a menos de 5,25																				de 5,25 a menos de 5,55																				de 5,55 a menos de 5,85																				de 5,85 a menos de 6,15																				de 6,15 a menos de 6,45																				de 6,45 a menos de 6,75																				de 6,75 a menos de 7,05																				de 7,05 a menos de 7,35																				de 7,35 a menos de 7,65																				de 7,65 a menos de 7,95																				de 7,95 a menos de 8,25																				de 8,25 a menos de 8,55																				de 8,55 a menos de 8,85																				de 8,85 a menos de 9,15																				de 9,15 a menos de 9,45																				de 9,45 a menos de 9,75																				de 9,75 a menos de 10,05																				de 10,05 a menos de 10,35																				FRECUENCIAS		$x - y$		Frecuencias Diag.		(x-y) Frecuencias Diag.		(x-y) ² Frecuencias Diag.	
		0,30	0,60	0,90	1,20	1,50	1,80	2,10	2,40	2,70	3,00	3,30	3,60	3,90	4,20	4,50	4,80	5,10	5,40	5,70	6,00	6,30	6,60	6,90	7,20	7,50	7,80	8,10	8,40	8,70	9,00	9,30	9,60	9,90	10,20	1	2	16,60	1	-16,60	275	50																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																									
De -6,45 a menos de -6,15	-6,30												1																							1	2																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
de -6,15 a menos de -5,85	-6,00																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -5,85 a menos de -5,55	-5,70																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -5,55 a menos de -5,25	-5,40																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -5,25 a menos de -4,95	-5,10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -4,95 a menos de -4,65	-4,80																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -4,65 a menos de -4,35	-4,50																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -4,35 a menos de -4,05	-4,20																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -4,05 a menos de -3,75	-3,90														1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
de -3,75 a menos de -3,45	-3,60																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -3,45 a menos de -3,15	-3,30																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -3,15 a menos de -2,85	-3,00																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -2,85 a menos de -2,55	-2,70																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -2,55 a menos de -2,25	-2,40																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -2,25 a menos de -1,95	-2,10																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -1,95 a menos de -1,65	-1,80																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -1,65 a menos de -1,35	-1,50	1																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
de -1,35 a menos de -1,05	-1,20																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -1,05 a menos de -0,75	-0,90																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
de -0,75 a menos de -0,45	-0,60																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																		
FRECUENCIAS		1	—	—	4	4	9	14	13	7	10	7	3	2	3	1	1	1	1	—	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	83																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
$x - y$		-0,80	-1,20	-1,50	-1,80	-2,10	-2,40	-2,70	-3,00	-3,30	-3,60	-3,90	-4,20	-4,50	-4,80	-5,10	-5,40	-5,70	-6,00	-6,30	-6,60	-6,90	-7,20	-7,50	-7,80	-8,10	-8,40	-8,70	-9,00	-9,30	-9,60	-9,90	-10,20	-10,50																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
Frecuencias Diag.		—	—	—	2	—	—	3	1	2	6	11	9	8	1	8	6	5	5	2	2	3	1	2	—	1	—	—	1	—	—	2	1	—																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
(x-y) Frecuencias Diag.		—	—	—	-3,60	—	—	-8,10	-3,00	-6,60	-21,60	-42,90	-37,80	-36,00	-4,80	-40,80	-32,40	-28,50	-30,00	-12,60	-13,20	-20,70	-7,20	-15,00	—	-8,10	—	—	-9,00	—	—	-19,80	-10,20	—																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
(x-y) ² Frecuencias Diag.		—	—	—	6,48	—	—	21,87	9,00	21,78	77,76	167,31	158,76	162,00	23,04	208,08	174,96	162,49	180,00	79,38	87,12	142,83	51,84	112,50	—	65,61	—	—	81,00	—	—	196,02	104,04	—																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	

<div><div>$y_{134} =$ $= -y_1 + y_3 - y_4$</div><div>$x_{134} =$ $= x_1 - x_3 - x_4$</div></div>	De -0,15 a menos de 0,15	de 0,15 a menos de 0,45	de 0,45 a menos de 0,75	de 0,75 a menos de 1,05	de 1,05 a menos de 1,35	de 1,35 a menos de 1,65	de 1,65 a menos de 1,95	de 1,95 a menos de 2,25	de 2,25 a menos de 2,55	de 2,55 a menos de 2,85	de 2,85 a menos de 3,15	de 3,15 a menos de 3,45	de 3,45 a menos de 3,75	de 3,75 a menos de 4,05	de 4,05 a menos de 4,35	de 4,35 a menos de 4,65	de 4,65 a menos de 4,95	de 4,95 a menos de 5,25	de 5,25 a menos de 5,55	de 5,55 a menos de 5,85	de 5,85 a menos de 6,15	de 6,15 a menos de 6,45	de 6,45 a menos de 6,75	de 6,75 a menos de 7,05	de 7,05 a menos de 7,35	de 7,35 a menos de 7,65	de 7,65 a menos de 7,95	de 7,95 a menos de 8,25	de 8,25 a menos de 8,55	de 8,55 a menos de 8,85	de 8,85 a menos de 9,15	de 9,15 a menos de 9,45	de 9,45 a menos de 9,75	FRECUENCIAS	<div><div>$x - y$</div><div>Frecuencias Diag.</div><div>(x-y) Frecuencias Diag.</div><div>(x-y)² Frecuencias Diag.</div></div>														
	0	0,30	0,60	0,90	1,20	1,50	1,80	2,10	2,40	2,70	3,00	3,30	3,60	3,90	4,20	4,50	4,80	5,10	5,40	5,70	6,00	6,30	6,60	6,90	7,20	7,50	7,80	8,10	8,40	8,70	9,00	9,30	9,60																
De -5,55 a menos de -5,25	-5,40																																		1	1	-15,00	1	-15,00	225,00									
de -5,25 a menos de -4,95	-5,10																																				1		-14,70										
de -4,95 a menos de -4,65	-4,80										1																											1		-14,40									
de -4,65 a menos de -4,35	-4,50																																						1		-14,10								
de -4,35 a menos de -4,05	-4,20																																						1		-13,80								
de -4,05 a menos de -3,75	-3,90												1																										1		-13,50								
de -3,75 a menos de -3,45	-3,60																																							1		-13,20							
de -3,45 a menos de -3,15	-3,30																																							1		-12,90							
de -3,15 a menos de -2,85	-3,00										1																												3		-12,60								
de -2,85 a menos de -2,55	-2,70											1			1				1																				2		-12,30								
de -2,55 a menos de -2,25	-2,40						1									1																							8		-12,00								
de -2,25 a menos de -1,95	-2,10					1		2	1	1	3																												15		-11,70								
de -1,95 a menos de -1,65	-1,80						1	1	3	4	3	3																											14		-11,40								
de -1,65 a menos de -1,35	-1,50					1	3	3	3	2					1	1																								20		-11,10							
de -1,35 a menos de -1,05	-1,20				2	5	3	6	3										1																					7		-10,80							
de -1,05 a menos de -0,75	-0,90	1			1	1	1	1		1																														3		-10,50							
de -0,75 a menos de -0,45	-0,60				1					1																															4		-10,20						
de -0,45 a menos de -0,15	-0,30				3			1																																	1		-9,90						
de -0,15 a menos de 0,15	0				1																																					1		-9,60					
de 0,15 a menos de 0,45	0,30	1																																								1		-9,30					
de 0,45 a menos de 0,75	0,60																																											1		-9,00			
de 0,75 a menos de 1,05	0,90																																											1		-8,70			
de 1,05 a menos de 1,35	1,20																																											1		-8,40			
de 1,35 a menos de 1,65	1,50										1																																	1		-8,10			
FRECUENCIAS		2	—	6	4	8	8	13	12	8	6	8	—	1	2	2	—	—	1	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	83												
$x - y$		1,50	1,20	0,90	0,60	0,30	0	-0,30	-0,60	-0,90	-1,20	-1,50	-1,80	-2,10	-2,40	-2,70	-3,00	-3,30	-3,60	-3,90	-4,20	-4,50	-4,80	-5,10	-5,40	-5,70	-6,00	-6,30	-6,60	-6,90	-7,20	-7,50	-7,80																
Frecuencias Diag.		—	—	—	—	1	—	—	1	4	2	2	1	4	6	6	9	9	4	5	7	4	4	3	1	3	—	—	3	—	—	—	1	2															
(x-y) Frecuencias Diag.		—	—	—	—	0,30	—	—	-0,60	-3,60	-2,40	-3,00	-1,80	-8,40	-14,40	-16,20	-27,00	-29,70	-14,40	-19,50	-29,40	-18,00	-19,20	-15,30	-5,40	-17,10	—	—	-19,80	—	—	-7,50	-15,60																
(x-y) ² Frecuencias Diag.		—	—	—	—	0,09	—	—	0,36	3,24	2,88	4,50	3,24	17,64	34,56	42,12	81,00	98,01	51,84	76,05	123,48	81,00	92,16	78,03	29,16	97,47	—	—	130,68	—	—	56,25	121,68																

<div><div>$y_{234} =$ $= y_2 + y_3 + y_4$</div><div>$x_{234} =$ $= -x_2 - x_3 - x_4$</div></div>	De 4,75 a menos de 5,25	de 5,25 a menos de 5,75	de 5,75 a menos de 6,25	de 6,25 a menos de 6,75	de 6,75 a menos de 7,25	de 7,25 a menos de 7,75	de 7,75 a menos de 8,25	de 8,25 a menos de 8,75	de 8,75 a menos de 9,25	de 9,25 a menos de 9,75	de 9,75 a menos de 10,25	de 10,25 a menos de 10,75	de 10,75 a menos de 11,25	de 11,25 a menos de 11,75	de 11,75 a menos de 12,25	de 12,25 a menos de 12,75	de 12,75 a menos de 13,25	de 13,25 a menos de 13,75	de 13,75 a menos de 14,25	de 14,25 a menos de 14,75	de 14,75 a menos de 15,25	de 15,25 a menos de 15,75	de 15,75 a menos de 16,25	de 16,25 a menos de 16,75	de 16,75 a menos de 17,25	de 17,25 a menos de 17,75	de 17,75 a menos de 18,25	de 18,25 a menos de 18,75	de 18,75 a menos de 19,25	de 19,25 a menos de 19,75	de 19,75 a menos de 20,25	de 20,25 a menos de 20,75	de 20,75 a menos de 21,25	FRECUENCIAS	$x - y$	Frecuencias Diag.	$(x-y) \times$ Frecuencias Diag.	$(x-y)^2 \times$ Frecuencias Diag.	
	5,00	5,50	6,00	6,50	7,00	7,50	8,00	8,50	9,00	9,50	10,00	10,50	11,00	11,50	12,00	12,50	13,00	13,50	14,00	14,50	15,00	15,50	16,00	16,50	17,00	17,50	18,00	18,50	19,00	19,50	20,00	20,50							
De -19,25 a menos de -18,75	-19,00																			1															1	-39,50	—	—	—
de -18,75 a menos de -18,25	-18,50																																	—	-39,00	—	—	—	
de -18,25 a menos de -17,75	-18,00																																	—	-38,50	—	—	—	
de -17,75 a menos de -17,25	-17,50																																	—	-38,00	—	—	—	
de -17,25 a menos de -16,75	-17,00																																	—	-37,50	—	—	—	
de -16,75 a menos de -16,25	-16,50																			1													1	-37,00	—	—	—		
de -16,25 a menos de -15,75	-16,00																																	—	-36,50	—	—	—	
de -15,75 a menos de -15,25	-15,50																						1					1						2	-36,00	—	—	—	
de -15,25 a menos de -14,75	-15,00																																1	1	-35,50	1	-35,50	1 260,25	
de -14,75 a menos de -14,25	-14,50																				1												1	-35,00	—	—	—		
de -14,25 a menos de -13,75	-14,00																																	2	-34,50	—	—	—	
de -13,75 a menos de -13,25	-13,50																			1		2											3	-34,00	1	34,00	1 156,00		
de -13,25 a menos de -12,75	-13,00																1				2			1									4	-33,50	1	-33,50	1 122,25		
de -12,75 a menos de -12,25	-12,50															1		1	1														3	-33,00	—	—	—		
de -12,25 a menos de -11,75	-12,00											1					3	1															5	-32,50	—	—	—		
de -11,75 a menos de -11,25	-11,50								1					2	1	2	1	1															8	-32,00	—	—	—		
de -11,25 a menos de -10,75	-11,00												1		2	2	1				1												8	-31,50	1	-31,50	992,25		
de -10,75 a menos de -10,25	-10,50									3		1	2	4		3																		13	-31,00	1	-31,00	961,00	
de -10,25 a menos de -9,75	-10,00	1						1	1	1		5	3																					12	-30,50	—	—	—	
de -9,75 a menos de -9,25	-9,50							1			2		1			1																		6	-30,00	1	-30,00	900,00	
de -9,25 a menos de -8,75	-9,00								3			1																						4	-29,50	2	-59,00	1 740,50	
de -8,75 a menos de -8,25	-8,50					1			1	1						1																		4	-29,00	—	—	—	
de -8,25 a menos de -7,75	-8,00							1																										2	-28,50	4	-114,00	3 249,00	
de -7,75 a menos de -7,25	-7,50												1																					1	-28,00	2	-56,00	1 568,00	
de -7,25 a menos de -6,75	-7,00								1																									1	-27,50	1	-27,50	756,25	
de -6,75 a menos de -6,25	-6,50																																	—	-27,00	—	—	—	
de -6,25 a menos de -5,75	-6,00		1																															1					
FRECUENCIAS		1	1	—	—	1	—	3	3	10	3	7	7	7	3	6	5	6	3	2	3	6	—	2	1	—	1	—	1	—	—	—	1	83					
$x - y$		-11,00	-11,50	-12,00	-12,50	-13,00	-13,50	-14,00	-14,50	-15,00	-15,50	-16,00	-16,50	-17,00	-17,50	-18,00	-18,50	-19,00	-19,50	-20,00	-20,50	-21,00	-21,50	-22,00	-22,50	-23,00	-23,50	-24,00	-24,50	-25,00	-25,50	-26,00	-26,50						
Frecuencias Diag.		—	1	—	—	—	—	—	—	1	2	1	—	1	3	4	3	4	4	6	5	3	4	1	4	4	3	3	1	5	1	3	1						
$(x-y) \times$ Frecuencias Diag.		—	-11,50	—	—	—	—	—	—	-15,00	-31,00	-16,00	—	-17,00	-52,50	-72,00	-55,50	-76,00	-78,00	-120,00	-102,50	-63,00	-86,00	-22,00	-90,00	-92,00	-70,50	-72,00	-24,50	-125,00	-25,50	-78,00	-26,50						
$(x-y)^2 \times$ Frecuencias Diag.		—	132,25	—	—	—	—	—	—	225,00	480,50	256,00	—	289,00	918,75	1 296,00	1 026,75	1 444,00	1 521,00	2 400,00	2 101,25	1 323,00	1 849,00	484,00	2 025,00	2 116,00	1 656,75	1 728,00	600,25	3 125,00	650,25	2 028,00	702,25						

$x_{1234} =$ $= x_1 - x_2 - x_3 - x_4$		$y_{1234} =$ $= y_1 + y_2 + y_3 + y_4$																												Frecuencias	x - y	Frecuencias Diag.	(x-y) Frecuencias Diag.	(x-y) ² Frecuencias Diag.				
		De 2,20 a menos de 2,60	de 2,60 a menos de 3,00	de 3,00 a menos de 3,40	de 3,40 a menos de 3,80	de 3,80 a menos de 4,20	de 4,20 a menos de 4,60	de 4,60 a menos de 5,00	de 5,00 a menos de 5,40	de 5,40 a menos de 5,80	de 5,80 a menos de 6,20	de 6,20 a menos de 6,60	de 6,60 a menos de 7,00	de 7,00 a menos de 7,40	de 7,40 a menos de 7,80	de 7,80 a menos de 8,20	de 8,20 a menos de 8,60	de 8,60 a menos de 9,00	de 9,00 a menos de 9,40	de 9,40 a menos de 9,80	de 9,80 a menos de 10,20	de 10,20 a menos de 10,60	de 10,60 a menos de 11,00	de 11,00 a menos de 11,40	de 11,40 a menos de 11,80	de 11,80 a menos de 12,20	de 12,20 a menos de 12,60	de 12,60 a menos de 13,00	de 13,00 a menos de 13,40						de 13,40 a menos de 13,80	de 13,80 a menos de 14,20	de 14,20 a menos de 14,60	de 14,60 a menos de 15,00
		2,40	2,80	3,20	3,60	4,00	4,40	4,80	5,20	5,60	6,00	6,40	6,80	7,20	7,60	8,00	8,40	8,80	9,20	9,60	10,00	10,40	10,80	11,20	11,60	12,00	12,40	12,80	13,20	13,60	14,00	14,40	14,80					
De -11,80 a menos de -11,40	-11,60																1																	1	26,40	—	—	—
de -11,40 a menos de -11,00	-11,20																																	—	-26,00	—	—	—
de -11,00 a menos de -10,60	-10,80																																	—	-25,60	—	—	—
de -10,60 a menos de -10,20	-10,40																		1														1	-25,20	—	—	—	
de -10,20 a menos de -9,80	-10,00																																1	-24,80	1	-24,80	615,04	
de -9,80 a menos de -9,40	-9,60																																	—	-24,40	—	—	—
de -9,40 a menos de -9,00	-9,20																									1							1	-24,00	—	—	—	
de -9,00 a menos de -8,60	-8,80																							1									1	-23,60	—	—	—	
de -8,60 a menos de -8,20	-8,40																	1		1												2	-23,20	—	—	—		
de -8,20 a menos de -7,80	-8,00															1		1		1												3	-22,80	—	—	—		
de -7,80 a menos de -7,40	-7,60																1	1														2	-22,40	—	—	—		
de -7,40 a menos de -7,00	-7,20														2		3															5	-22,00	—	—	—		
de -7,00 a menos de -6,60	-6,80							1	1				1	2	1	1						1										8	-21,60	—	—	—		
de -6,60 a menos de -6,20	-6,40										1			2	1	1		1														6	-21,20	1	-21,20	449,44		
de -6,20 a menos de -5,80	-6,00							1	1	2	3	2	2					1			1											13	-20,80	—	—	—		
de -5,80 a menos de -5,40	-5,60						1	2	2	2	4	1	1																				13	-20,40	—	—	—	
de -5,40 a menos de -5,00	-5,20	1					2	4	3	3	1																						16	-20,00	1	-20,00	400,00	
de -5,00 a menos de -4,60	-4,80							2																									3	-19,60	2	-39,20	768,32	
de -4,60 a menos de -4,20	-4,40					2						1																					3	-19,20	—	—	—	
de -4,20 a menos de -3,80	-4,00																																—	-18,80	—	—	—	
de -3,80 a menos de -3,40	-3,60			1			2																										3	-18,40	—	—	—	
de -3,40 a menos de -3,00	-3,20																																—	-18,00	1	-18,00	324,00	
de -3,00 a menos de -2,60	-2,80																																—	-17,60	1	-17,60	309,76	
de -2,60 a menos de -2,20	-2,40										1																						1					
Frecuencias		1	—	1	—	—	5	5	10	7	7	11	4	7	4	3	5	5	1	2	1	1	1	—	—	1	—	—	—	—	—	—	1	83				
x - y		-4,80	-5,20	-5,60	-6,00	-6,40	-6,80	-7,20	-7,60	-8,00	-8,40	-8,80	-9,20	-9,60	-10,00	-10,40	-10,80	-11,20	-11,60	-12,00	-12,40	-12,80	-13,20	-13,60	-14,00	-14,40	-14,80	-15,20	-15,60	-16,00	-16,40	-16,80	-17,20					
Frecuencias Diag.		—	—	—	—	—	1	—	1	—	2	3	1	2	4	5	6	6	4	7	5	4	2	3	3	2	4	1	3	3	1	1	2					
(x-y) Frecuencias Diag.		—	—	—	—	—	-6,80	—	-7,60	—	-16,80	-26,40	-9,20	-19,20	-40,00	-52,00	-64,80	-67,20	-46,40	-84,00	-62,00	-51,20	-26,40	-40,80	-42,00	-28,80	-59,20	-15,20	-46,80	-48,00	-16,40	-16,80	-34,40					
(x-y) ² Frecuencias Diag.		—	—	—	—	—	46,24	—	57,76	—	141,12	232,32	84,64	184,32	400,00	540,80	699,84	752,64	538,24	1 008,00	768,80	655,36	348,48	554,88	588,00	414,72	876,16	231,04	730,08	768,00	268,96	282,24	591,68					

A P É N D I C E

ESTADISTICA INDUSTRIAL - AÑO 1956

Tabla A

CONCEPTOS	Obreros ocupados	Horas- Obrero Trabaja- das	Monto de los salarios pagados	Volumen físico de la produc. industrial	Pod. adquis. del monto sal. pagad. (3) × 0,7274
	(1)	(2)	(3)	(4)	(5)
RAMAS DE INDUSTRIA					
1) INDUSTRIA EXTRACTIVA					
Diversos minerales.	100,8	108,3	155,4	122,0	113,0
Minerales metalíferos.	123,9	136,2	223,7	122,2	162,7
Petróleo - Yacimientos.	107,9	146,6	206,2	124,9	150,0
Sal común yacimientos.	103,4	106,1	168,3	71,3	122,4
2) INDUSTRIA MANUFACTURERA					
a) ALIMENTOS Y BEBIDAS					
Aceites comestibles.	70,0	72,4	113,6	89,2	82,6
Aguas gaseosas y bebidas sin alcohol	98,3	106,2	168,7	157,6	122,7
Arroz.	149,4	151,0	271,1	112,4	197,2
Azúcar.	96,1	96,0	146,1	133,6	106,3
Carne conservada y embutidos.	111,6	118,7	196,4	119,0	142,9
Frigoríficos.	114,3	127,8	196,0	171,1	142,6
Cerveza.	94,6	95,0	128,2	90,6	93,3
Chocolates, carameles, etc.	89,1	91,3	152,1	101,1	110,6
Pastas alimenticias.	87,9	95,9	150,6	113,0	109,5
Frutas y legumbres secas y en con- servas.	117,7	138,9	222,5	132,8	161,8
Galletitas y bizcochos.	100,1	100,7	154,8	114,1	112,6
Harina de trigo.	107,9	114,0	163,4	117,0	118,9
Malta, levadura y polvo de hornear.	93,0	107,7	132,7	94,5	96,5
Pescado y mariscos en conservas.	109,3	96,5	167,7	86,4	122,0
Varios.	97,8	91,6	149,4	119,6	108,7
b) TABACO					
Cigarros y Cigarrillos.	85,3	87,1	140,1	97,9	101,
c) TEXTILES					
Hilados y Tejidos de lana, Algodón y otras fib.	90,6	97,4	146,9	113,7	106,9
Medias.	89,8	95,0	141,1	107,8	102,6
Tejidos de Punto, de lana, algodón, etc.	78,7	89,8	132,1	131,9	96,1
Tejidos de seda.	73,0	78,3	107,9	109,9	78,5
Tintorerías industriales.	84,0	87,2	155,4	101,7	113,0
Trenzas, Sogas, Cabos.	86,5	97,0	140,3	116,8	102,1
d) CONFECCIONES					
Bolsas de arpillera.	105,9	128,5	190,0	183,0	138,2
Alpargatas.	98,0	107,0	150,5	125,7	109,5
Grandes tiendas.	50,9	54,4	82,6	65,9	60,1
Roperías.	122,8	135,4	191,0	89,6	138,9
Sastrerías.	105,1	107,9	146,8	108,1	106,8
Sombreros de fieltro.	54,8	63,1	104,1	78,0	75,7
e) MADERAS					
Cajones para envase y embalaje.	83,6	92,8	135,1	94,5	98,3
Muebles.	84,1	85,2	133,2	78,8	96,9
Carpintería de obras.	69,8	66,6	99,0	143,7	72,0
Toneles, cascos, barricas, barriles, etc.	74,7	80,4	129,5	112,7	94,2
f) PAPEL Y CARTON					
Celulosa.	133,8	120,0	231,8	166,2	168,0
Papel, cartón y cartulina.	121,6	128,4	253,0	161,1	184,0
Sobres y bolsas de papel.	104,9	102,0	163,1	102,0	184,0
g) IMPRENTA Y PUBLICACIONES					
Diarios, periódicos y revistas.	103,7	100,7	186,7	192,6	135,8
Imprenta.	105,8	93,2	154,2	122,0	112,1
h) PRODUCTOS QUIMICOS					
Colores, pinturas y barnices.	86,9	92,6	145,4	131,3	105,8
Curtientes.	79,7	87,3	170,0	79,1	123,7
Fosforos.	94,1	102,7	176,7	103,7	128,5

ESTADISTICA INDUSTRIAL - AÑO 1956

Tabla A

CONCEPTOS	Obreros ocupados	Horas- Obrero Trabajadas	Monto de los salarios pagados	Volumen físico de la produc. industrial	Pod. adquis. del monto sal. pagad. (3) × 0,7274 (5)
	(1)	(2)	(3)	(4)	(5)
Gases comprimidos y licuados . . .	93,5	90,8	136,9	121,2	99,6
Rayón	123,0	140,3	200,0	163,0	145,5
Jabón común para lavar	101,3	107,6	176,9	124,6	128,7
Substancias y productos químicos . .	85,4	86,4	130,6	111,4	95,0
Tintas para imprenta	112,3	110,5	169,2	146,2	123,1
i) DERIVADOS DEL PETROLEO					
Refinerías de petróleo	114,5	117,7	206,2	123,1	150,0
j) CAUCHO					
Calzado de caucho	128,9	131,1	223,7	170,6	162,7
Cubiertas y cámaras para automotores	100,4	103,4	186,9	104,6	136,0
k) CUERO					
Calzado	90,1	97,4	161,5	93,8	117,5
Cueros y pieles; curtido, teñido y apresto	90,5	98,0	158,2	104,8	115,1
l) PIEDRAS, VIDRIO Y CERAMICA					
Alfarería y cerámica	103,3	107,3	176,2	115,2	128,2
Artículos de cemento y fibrocemento .	79,0	84,2	138,3	123,3	100,6
Cemento	108,6	112,6	175,3	132,4	127,5
Ladrillos de máquina	104,7	117,7	181,3	145,0	131,9
Mosaicos	71,3	77,8	111,3	85,8	81,0
Fábricas de vidrio	88,3	100,0	148,7	115,7	108,2
m) METALES (EXCLUIDA MAQUINARIA)					
Artículos diversos de hierro	106,5	96,8	153,3	124,1	111,5
Hojalaterías	105,0	105,8	162,8	144,4	118,4
Bulones, tornillos, etc.	75,1	73,2	113,5	93,4	82,6
Caños de hierro y acero	137,7	138,1	206,9	171,0	150,5
Fundición y elaboración de cobre y bronce	108,5	107,3	176,5	126,9	128,4
Cocinas y otros art. excluidos los eléctricos	104,4	100,3	156,4	147,5	113,8
Fund. y elabor. de hierro, acero y otros metales	99,8	96,0	141,3	139,1	102,8
Hierro, galvanización	148,2	134,9	222,9	197,7	162,1
Fund. y elab. de plomo, estaño, y otros metales no ferrosos	111,2	105,2	136,7	192,4	99,4
Fundición de minerales de plomo, estaño y otros metales no ferrosos .	96,1	105,3	160,3	98,2	116,6
Carpintería metálica	88,0	79,7	118,0	68,9	85,8
n) VEHICULOS Y MAQUINARIA (EXCLUIDA LA ELECTRICA)					
Astilleros y talleres navales	77,6	55,2	89,6	55,2	65,2
Armado de automotores y fabricac. de carrocerías	96,0	93,7	144,4	231,7	105,0
Maquinarias y motores, excluidos los eléctricos	104,1	98,4	150,5	98,4	109,5
ñ) MAQUINARIA Y APARATOS ELECTRICOS					
Acumuladores, pilas y baterías . . .	88,0	83,3	131,5	86,7	95,7
Artículos y aparatos diversos para electric.	128,7	127,4	224,3	130,7	163,2
Lámparas y tubos para electricidad .	127,2	120,0	222,3	140,8	161,7
Motores eléctricos	104,8	91,9	142,4	116,3	103,6
Radiofonía	136,0	126,9	222,0	153,7	161,5
o) VARIOS					
Botones, boquillas, peines, etc. . . .	82,3	88,2	144,3	84,7	105,0
Hielo	92,0	93,8	147,8	88,8	107,5
ELECTRICIDAD Y GAS					
Fábricas de electricidad	113,8	112,9	185,2	125,4	134,7
Gas manufacturado	106,7	104,0	157,3	129,6	114,4

ESTADISTICA INDUSTRIAL - AÑO 1955

Tabla B

CONCEPTOS	Obreros ocupados	Horas- Obrero Trabaja- das	Poder Adquisi- tivo del Monto de los Salarios Pagados (5) × 0,825	Volumen Físico de la Prod. Industr.	Monto de los Salarios Pagados
	(1)	(2)	(3)	(4)	(5)
RAMAS DE INDUSTRIA					
1) INDUSTRIA EXTRACTIVA					
Diversos minerales.	97,6	102,3	104,6	118,2	126,8
Minerales metalíferos.	118,1	129,0	144,9	115,2	175,6
Petróleo - Yacimientos.	106,5	142,9	153,2	122,8	185,7
Sal común - Yacimientos.	91,1	89,3	97,6	45,4	118,3
2) INDUSTRIA MANUFACTURERA					
a) ALIMENTOS Y BEBIDAS					
Aceites comestibles.	71,6	69,9	79,0	58,8	95,7
Aguas gaseosas y bebidas sin alcohol.	108,8	108,5	119,0	127,8	144,3
Arroz.	136,5	138,7	185,0	105,8	224,2
Azúcar.	99,6	102,1	114,8	110,0	139,2
Carne conservada y embutidos.	110,8	113,7	141,8	110,9	171,9
Frigoríficos.	106,3	112,4	127,3	136,7	154,3
Cerveza.	97,7	99,5	101,5	94,9	123,0
Chocolates, caramelos, etc.	92,7	92,4	103,8	100,8	125,8
Pastas alimenticias.	89,5	96,0	115,9	108,0	140,5
Frutas y legumbres secas y en conserva.	111,0	120,5	141,6	167,8	171,6
Galletitas y bizcochos.	92,1	90,6	103,5	100,6	125,4
Harina de trigo.	106,7	110,7	116,7	111,0	141,5
Malta, levadura y polvo de hornear.	98,9	108,0	100,7	85,8	122,1
Pescado y mariscos en conservas.	114,5	109,0	109,6	95,8	132,8
Varios.	100,3	90,5	26,2	107,5	31,8
b) TABACO					
Cigarros y Cigarrillos.	93,6	94,4	105,9	100,2	128,4
c) TEXTILES					
Hilados y Tejidos de lana, Algodón y otras fíb.	91,7	95,5	102,1	112,3	123,8
Medias.	88,1	92,0	100,7	105,5	122,1
Tejidos de Punto, de lana, algodón, etc.	86,2	95,1	101,5	126,4	123,0
Tejidos de Seda.	74,5	77,9	80,5	103,5	97,6
Tintorerías Industriales.	94,8	96,0	112,8	106,3	136,7
Trenzas, Sogas, Cabos.	87,7	97,1	97,1	107,6	117,7
d) CONFECCIONES					
Bolsas de arpillera.	124,2	147,9	159,4	212,2	193,2
Alpargatas.	99,1	107,6	108,7	121,4	131,8
Grandes tiendas.	54,5	56,9	61,6	72,1	74,7
Roperías.	128,3	137,2	133,0	103,1	161,2
Sastrerías.	104,6	106,0	106,8	104,6	129,5
Sombreros de fieltro.	69,8	79,7	96,9	91,8	117,5
e) MADERAS					
Cajones para envase y embalaje.	86,7	95,9	104,0	102,5	126,0
Muebles.	83,6	84,7	97,6	87,6	118,3
Carpintería de Obras.	78,0	79,3	85,7	125,7	103,9
Toneles, cascos, barricas, barriles, etc.	76,5	91,4	98,6	102,7	119,5
f) PAPEL Y CARTON					
Celulosa.	110,5	102,8	110,9	137,4	134,4
Papel, cartón y cartulina.	108,3	118,9	147,6	142,7	178,9
Sobre y bolsas de papel.	103,6	108,1	114,5	102,0	138,8
g) IMPRENTA Y PUBLICACIONES					
Diarios, periódicos y revistas.	110,2	103,5	118,0	124,7	143,0
Imprenta.	100,9	102,9	114,5	139,5	138,8
h) PRODUCTOS QUIMICOS					
Colores, pinturas y barnices.	93,8	96,6	106,6	119,6	129,2
Curtientes.	92,5	100,7	142,6	105,0	172,8
Fosforos.	96,0	99,6	108,3	96,4	131,3

ESTADISTICA INDUSTRIAL - AÑO 1955

Tabla B

CONCEPTOS	Obreros ocupados	Horas- Obrero Trabaja- das	Poder Adquisi- tivo del Monto de los Salarios Pagados (5) × 0,825	Volumen Físico de la Prod. Industr.	Monto de los Salarios Pagados
	(1)	(2)	(3)	(4)	(5)
Gases comprimidos y licuados	94,9	89,2	98,7	119,8	119,6
Rayón	121,7	132,1	133,9	150,2	162,3
Jabón común para lavar	102,8	109,4	122,4	115,8	148,4
Substancias y productos químicos	85,6	86,0	95,8	105,6	116,1
Tintas para imprenta	105,4	108,9	116,8	141,5	141,6
i) DERIVADOS DEL PETROLEO					
Refinerías de petróleo	109,7	114,7	144,8	119,0	175,5
j) CAUCHO					
Calzado de caucho	123,9	125,5	146,5	156,8	177,6
Cubiertas y cámaras para automotores	107,3	107,4	129,2	111,4	156,6
k) CUERO					
Calzado	96,3	100,1	120,3	96,1	145,8
Cueros y pieles; cuero curtido, teñido y apresto	91,3	96,6	110,7	103,4	134,2
l) PIEDRAS, VIDRIO Y CERAMICA					
Alfarería y cerámica	102,0	104,5	121,2	115,7	146,9
Artículos de cemento y fibrocemento	85,9	89,0	105,5	132,0	127,9
Cemento	104,1	105,7	110,7	119,6	134,2
Ladrillos de máquina	100,2	106,6	112,8	125,7	136,7
Mosaicos	74,2	79,6	83,0	86,8	100,6
Fábricas de vidrio	88,5	97,7	96,8	110,6	117,3
m) METALES (EXCLUIDA MAQUINARIA)					
Artículos diversos de hierro	103,7	100,0	115,2	136,8	139,6
Hojalaterías	103,7	110,3	120,0	151,3	145,5
Bulones, tornillos, etc.	78,8	83,8	92,7	101,3	112,4
Caños de hierro y acero	133,5	143,6	160,6	179,9	194,7
Fundición y elaboración de cobre y bronce	107,6	107,4	126,0	127,9	152,7
Cocinas y otros artef. excluidos los eléctricos	111,7	117,5	127,0	142,2	153,9
Fund. y elabor. de hierro, acero y otros metales	100,5	104,2	116,6	146,0	141,3
Hierro, galvanización	153,3	155,5	173,3	296,5	210,0
Fund. y elab. de plomo, estaño, y otros metales no ferrosos	107,9	114,3	119,5	185,6	144,9
Fundición de minerales de plomo, estaño y otros metales no ferrosos	102,3	109,5	122,0	86,7	147,9
Carpintería metálica	89,7	88,2	96,6	70,0	117,1
n) VEHICULOS Y MAQUINARIA (EXCLUIDA LA ELECTRICA)					
Astilleros y talleres navales	102,4	96,6	113,8	96,8	137,9
Armado de automotores y fabricac. de carrocerías	102,9	104,3	115,9	289,9	140,5
Maquinaria y motores, excluido los eléctricos	102,8	103,3	115,9	103,3	140,5
ñ) MAQUINARIA Y APARATOS ELECTRICOS					
Acumuladores, pilas y baterías	87,1	85,5	97,2	81,9	117,8
Artículos y aparatos diversos para electric.	127,2	128,9	152,2	133,1	184,5
Lámparas y tubos para electricidad	126,7	121,2	150,2	121,2	182,1
Motores eléctricos	108,0	107,3	127,5	140,0	154,5
Radiofonía	144,4	144,5	171,4	156,0	207,7
o) VARIOS					
Botones, boquillas, peines, etc.	88,9	90,2	101,7	95,2	123,3
Hielo	93,3	92,9	106,7	93,0	129,3
ELECTRICIDAD Y GAS					
Fábricas de electricidad	108,6	106,8	138,6	121,7	168,0
Gas manufacturado	103,5	98,1	112,4	112,6	136,3

ESTADISTICA INDUSTRIAL - AÑO 1957

Tabla C

CONCEPTOS	Obreros ocupados	Horas- Obrero Trabaja- das	Volumen Físico de la Produc. Industrial	Montó de los Salarios Pagados	Poder Adquisi- tivo de los Salarios Pagados (4) × 0,583
	(1)	(2)	(3)	(4)	(5)
RAMAS DE INDUSTRIA					
1) INDUSTRIA EXTRACTIVA					
Diversos minerales	94,1	93,8	154,0	167,9	97,9
Minerales metálicos	108,4	122,1	133,2	232,3	135,4
Petróleo - Yacimientos	112,0	146,6	136,7	351,7	205,0
Sal común - Yacimientos	102,4	106,4	61,8	189,4	110,4
2) INDUSTRIA MANUFACTURERA					
a) ALIMENTOS Y BEBIDAS					
Aceites comestibles	73,2	74,3	102,6	154,7	90,2
Aguas gaseosas y bebidas sin alcohol	96,5	107,5	190,5	224,1	130,7
Arroz	142,9	140,3	108,2	345,2	201,3
Azúcar	94,5	92,6	113,7	198,6	115,8
Carne conservada y embutidos	114,0	120,7	131,6	260,5	151,9
Frigoríficos	116,5	118,4	139,7	245,8	143,3
Cerveza	95,9	95,1	91,6	186,7	108,8
Chocolates, caramelos, etc.	90,8	93,1	106,6	201,2	117,3
Pastas alimenticias	88,0	95,2	115,7	201,9	117,7
Frutas y legumbres secas y en con- serva	128,0	132,7	144,9	307,4	179,2
Galletitas y bizcochos	108,7	106,2	133,6	201,9	117,7
Harina de trigo	105,0	110,6	116,8	235,5	137,3
Malta, levadura y polvo de hornear	96,4	106,5	92,8	168,6	98,3
Pescado y mariscos en conservas	95,2	82,9	91,8	168,1	98,0
b) TABACO					
Cigarros y Cigarillos	70,9	69,9	97,6	147,5	86,0
c) TEXTILES					
Hilados y Tejidos de lana, Algodón y otras fib.	88,0	91,4	104,7	180,3	105,1
Medias	89,3	89,4	105,9	173,4	101,1
Tejidos de punto, de lana, algodón, etc.	73,6	79,1	123,3	150,8	87,9
Tejidos de seda	76,5	81,2	114,5	148,5	86,6
Tintorerías Industriales	72,9	76,8	91,8	160,1	93,3
Trenzas, Sogas, Cabos	84,3	92,3	116,7	174,4	101,7
d) CONFECCIONES					
Bolsas de arpillera	106,2	129,1	189,0	237,3	138,3
Alpargatas	100,3	104,8	116,1	195,4	113,9
Grandes tiendas	47,1	47,4	60,4	104,0	60,6
Roperías	124,8	136,7	78,1	255,5	149,0
Sastrerías	107,1	100,9	98,7	192,2	112,1
Sombreros de fieltro	53,3	59,6	75,9	126,3	73,6
e) MADERAS					
Cajones para envase y embalaje	82,5	90,0	83,2	176,0	102,6
Muebles	81,9	85,8	94,6	167,4	97,6
Carpintería de Obras	69,3	67,6	145,6	133,1	77,6
Toneles, cascos, barricas, barriles, etc.	77,4	79,7	98,7	175,4	102,3
f) PAPEL Y CARTON					
Celulosa	179,2	177,2	174,2	308,0	179,6
Papel, cartón y cartulina	124,8	131,0	161,0	284,1	165,6
Sobre y bolsas de papel	112,0	115,5	115,0	237,8	138,6
g) IMPRENTA Y PUBLICACIONES					
Diarios, periódicos y revistas	106,4	107,0	249,6	254,8	148,5
Imprenta	105,6	108,5	142,0	242,6	141,4
h) PRODUCTOS QUIMICOS					
Colores, pinturas y barnices	87,2	92,0	133,9	193,3	112,7
Curtientes	78,3	85,1	80,5	194,7	113,5
Fosforos	86,7	89,6	96,6	205,1	119,6

ESTADISTICA INDUSTRIAL - AÑO 1957

Tabla C

CONCEPTOS	Obreros ocupados	Horas- Obrero Trabaja- das	Volumen Físico de la Produc. Industrial	Monto de los Salarios Pagados	Poder Adquisi- tivo de los Salarios Pagados (4) × 0,583
	(1)	(2)	(3)	(4)	(5)
Gases comprimidos y licuados . . .	93,4	90,9	128,1	193,9	113,0
Rayón	121,3	131,6	181,4	260,0	151,6
Jabón común para lavar	103,2	108,0	127,2	242,8	141,6
Substancias y productos químicos . .	88,5	87,0	109,5	184,9	107,8
Tintas para imprenta	110,3	108,7	162,6	222,5	129,7
i) DERIVADOS DEL PETROLEO					
Refinerías de petróleo	118,6	114,7	149,3	295,2	172,1
j) CAUCHO					
Calzado de caucho	126,8	127,9	169,8	276,3	161,1
Cubiertas y cámaras para automotores	101,7	102,8	111,8	229,1	133,6
k) CUERO					
Calzado	91,3	95,4	91,0	188,7	110,0
Cueros y pieles; curtido, teñido y apresto	89,7	98,3	108,4	203,4	118,6
l) PIEDRAS, VIDRIO Y CERAMICA					
Alfarería y cerámica	106,7	109,9	126,0	229,7	133,9
Artículos de cemento y fibrocemento .	71,8	76,2	108,1	161,5	94,2
Cemento	115,5	118,7	152,5	222,8	129,9
Ladrillos de máquina	104,9	114,2	145,0	213,8	124,6
Mosaicos	70,1	75,6	84,5	142,5	83,1
Fábricas de vidrio	88,8	98,9	123,0	185,3	108,0
m) METALES (EXCLUIDA MAQUINARIA)					
Artículos diversos de hierro	109,3	110,8	141,5	237,9	138,7
Hojalaterías	109,6	114,6	154,9	237,0	138,2
Bulones, tornillos, etc.	72,6	79,0	103,9	165,9	96,7
Caños de hierro y acero	136,1	154,5	257,1	321,2	187,3
Fundición y elaboración de cobre y bronce	104,1	108,5	167,1	231,5	135,0
Cocinas y otros artf. excluidos los eléctricos	111,9	118,7	171,3	244,7	142,7
Fund. y elabor. de hierro, acero y otros metales	97,0	101,3	138,9	215,9	125,9
Hierro, galvanización	121,0	121,0	180,0	259,1	151,1
Fund. y elab. de plomo, estaño, y otros metales no ferrosos	115,5	126,4	221,5	242,1	141,1
Fundición de minerales de plomo, estaño y otros metales no ferrosos .	100,4	108,2	110,8	250,7	146,2
Carpintería metálica	83,0	83,1	75,1	164,1	95,7
n) VEHICULOS Y MAQUINARIA (EXCLUIDA LA ELECTRICA)					
Astilleros y talleres navales	59,9	48,7	48,7	102,7	59,9
Armado de automotores y fabricac. de carrocerías	112,6	118,4	458,2	246,7	143,8
Maquinaria y motores, excluido los eléctricos	104,3	104,6	104,6	207,1	120,7
ñ) MAQUINARIA Y APARATOS ELECTRICOS					
Acumuladores, pilas y baterías . . .	93,3	95,1	122,0	203,6	118,7
Artículos y aparatos diversos para eléctric.	119,9	125,7	147,4	275,5	160,6
Lámparas y tubos para electricidad .	129,0	137,4	179,8	342,2	199,5
Motores eléctricos	102,0	95,9	76,9	231,5	135,0
Radiofonía	133,8	149,9	171,8	298,8	174,2
o) VARIOS					
Botones, boquillas, peines, etc. . . .	77,0	78,0	75,7	162,0	94,4
Hielo	93,1	92,9	81,1	182,3	106,3
ELECTRICIDAD Y GAS					
Fábricas de electricidad	118,4	117,0	130,9	289,4	168,7
Gas manufacturado	114,7	113,1	120,4	223,4	130,2

ESTADISTICA INDUSTRIAL - AÑO 1955-1957

$$\text{VALORES REDUCIDOS} = \frac{\text{Valor Real}}{\text{Diferencia Media}}$$

FACTOR DE REDUCCION: Diferencia Media entre cada serie del Año 1955 con su correspondiente del Año 1957

Tabla D

CONCEPTOS	Año 1955				Año 1957			
	Obre- ros ocu- pados	Horas- Obrero Traba- jadas	Poder adqui- sitivo del monto de los salarios pa- gados	Volu- men físico de la produc- ción in- dustrial	Obre- ros ocu- pados	Horas- Obrero Traba- jadas	Poder adqui- sitivo del monto de los salarios pa- gados	Volu- men físico de la produc- ción in- dustrial
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
RAMAS DE INDUSTRIA								
1) INDUSTRIA EXTRACTIVA								
Diversos minerales	4,73	4,51	3,33	2,69	4,56	4,19	3,12	3,51
Minerales metalíferos	5,72	5,69	4,61	2,62	5,25	5,43	4,31	3,03
Petróleo - Yacimientos	5,16	6,30	4,88	2,80	5,43	6,51	6,53	3,11
Sal común - Yacimientos	4,41	3,94	3,11	1,03	4,96	4,74	3,52	1,41
2) INDUS. MANUFACTURERA								
a) ALIMENTOS Y BEBIDAS								
Aceites comestibles	3,47	3,08	2,52	1,34	3,55	3,33	2,87	2,34
Aguas gaseosas y bebidas sin alcohol	5,27	4,78	3,79	2,91	4,68	4,79	4,16	4,34
Arroz	6,61	6,12	5,89	2,41	6,92	6,24	6,41	2,46
Azúcar	4,83	4,50	3,66	2,51	4,58	4,13	3,69	2,59
Carne conservada y embutidos	5,37	5,01	4,52	2,53	5,52	5,37	4,84	3,00
Frigoríficos	5,15	4,96	4,05	3,11	5,64	5,27	4,56	3,18
Cerveza	4,73	4,39	3,23	2,16	4,65	4,24	3,46	2,09
Chocolates, caramelos, etc.	4,49	4,07	3,31	2,30	4,40	4,15	3,74	2,43
Pastas alimenticias	4,34	4,23	3,69	2,46	4,26	4,25	3,75	2,64
Frutas y legumbres secas y en conserva	5,38	5,31	4,51	3,82	6,20	5,90	5,71	3,30
Galletitas y bizcochos	4,46	3,99	3,30	2,29	5,27	4,73	3,75	3,04
Harina de trigo	5,17	4,88	3,72	2,53	5,09	4,93	4,37	2,66
Malta, levadura y polvo de hornear	4,79	4,76	3,21	1,95	4,67	4,75	3,13	2,11
Pescado y mariscos en conserva	5,55	4,81	3,49	2,18	4,61	3,71	3,12	2,09
Varios	4,86	3,99	0,83	2,45	4,63	3,78	4,02	3,28
b) TABACO								
Cigarros y cigarrillos	4,53	4,16	3,37	2,28	3,44	3,13	2,74	2,22
c) TEXTILES								
Hilados y tejidos de lan, algodón y otras fib	4,44	4,21	3,25	2,56	4,26	4,08	3,35	2,38
Medias	4,27	4,06	3,21	2,40	4,33	3,99	3,22	2,41
Tejidos de punto, de lana, algodón, etc.	4,18	4,19	3,23	2,88	3,57	3,54	2,80	2,81
Tejidos de seda	3,61	3,43	2,56	2,36	3,71	3,63	2,76	2,61
Tintorerías industriales	4,59	4,23	3,59	2,42	3,53	3,43	2,97	2,09
Trenzas, sogas, cabos	4,25	4,28	3,09	2,45	4,08	4,12	3,24	2,66
d) CONFECCIONES								
Bolsas de arpillera	6,02	6,52	5,08	4,83	5,15	5,74	4,40	4,31
Alpargatas	4,80	4,74	3,46	2,77	4,86	4,67	3,63	2,64
Grandes tiendas	2,64	2,51	1,96	1,64	2,28	2,14	1,93	1,38

ESTADISTICA INDUSTRIAL - AÑO 1955-1957

VALORES REDUCIDOS = $\frac{\text{Valor Real}}{\text{Diferencia Media}}$

FACTOR DE REDUCCION: Diferencia Media entre cada serie del Año 1955 con su correspondiente del Año 1957

Tabla D

CONCEPTOS	AÑO 1955				AÑO 1957			
	Obre- ros ocu- pados	Horas- Obrero Traba- jadas	Poder adqui- sitivo del monto de los salarios pa- gados	Volu- men físico de la produc- ción in- dustrial	Obre- ros ocu- pados	Horas- Obrero Traba- jadas	Poder adqui- sitivo del monto de los salarios pa- gados	Volu- men físico de la produc- ción in- dustrial
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Roperías	6,22	6,05	4,23	2,35	6,05	6,08	4,75	1,78
Sastrerías	5,07	4,67	3,40	2,38	5,19	4,50	3,57	2,25
Sombreros de fieltro	3,38	3,51	3,09	2,09	2,58	2,63	2,34	1,73
e) MADERAS								
Cajones para envase y em- balaje	4,20	4,23	3,31	2,33	4,00	4,02	3,27	1,90
Muebles	4,05	3,73	3,11	2,00	3,97	3,83	3,11	2,15
Carpintería de obras	3,78	3,50	2,73	2,86	3,36	3,03	2,47	3,32
Toneles, cascos, barricas, bar- riles, etc.	3,71	4,03	3,14	2,34	3,75	3,56	3,26	2,25
f) PAPEL Y CARTON								
Celulosa	5,35	4,53	3,53	3,13	8,68	7,86	5,72	3,97
Papel, cartón y cartulina	5,25	5,24	4,70	3,25	6,05	5,83	5,27	3,67
Sobre y bolsas de papel	5,02	4,77	3,65	2,32	5,43	5,14	4,41	2,62
g) IMPRENTA Y PUBLICACIONES								
Diarios, periódicos y revistas	5,34	4,56	3,76	2,84	5,16	4,77	4,73	5,69
Imprenta	4,89	4,54	3,65	3,18	5,12	4,83	4,50	3,23
h) PRODUCTOS QUIMICOS								
Colores, pinturas y barnices	4,45	4,26	3,39	2,72	4,22	4,11	3,59	3,05
Curtientes	4,48	4,44	4,54	2,39	3,79	3,80	3,61	1,83
Fósforos	4,65	4,39	3,45	2,20	4,20	4,00	3,81	2,20
Gases comprimidos y licuados	4,60	3,93	3,14	2,73	4,53	4,06	3,60	2,92
Rayón	5,90	5,82	4,26	3,42	5,88	5,85	4,83	4,13
Jabón común para lavar	4,98	4,82	3,90	2,64	5,00	4,81	4,51	2,90
Substancias y productos quí- micos	4,15	3,79	3,05	2,41	4,29	3,89	3,43	2,49
Tintas para imprenta	5,11	4,80	3,72	3,22	5,34	4,84	4,13	3,70
i) DERIVADOS DEL PETROLEO								
Refinerías de petróleo	5,31	5,06	4,61	2,71	5,75	5,11	5,48	3,40
j) CAUCHO								
Calzado de caucho	6,00	5,53	4,67	3,57	6,14	5,69	5,13	3,87
Cubiertas y cámaras para auto- motores	5,20	4,74	4,11	2,54	4,93	4,58	4,25	2,55
k) CUERO								
Calzado	4,67	4,41	3,83	2,19	4,42	4,26	3,50	2,07
Cueros y pieles; curtido, te- ñido y apresto	4,42	4,26	3,53	2,36	4,35	4,38	3,78	2,47
l) PIEDRAS, VIDRIO Y CERAMICA								
Alfarería y cerámica	4,94	4,61	3,86	2,64	5,17	4,90	4,26	2,87
Artículos de cemento y fibro- cemento	4,16	3,92	3,36	3,01	3,48	3,41	3,00	2,46
Cemento	5,04	4,66	3,53	2,72	5,60	5,28	4,14	3,47

ESTADISTICA INDUSTRIAL - AÑO 1955-1957

$$\text{VALORES REDUCIDOS} = \frac{\text{Valor Real}}{\text{Diferencia Media}}$$

FACTOR DE REDUCCION : Diferencia Media entre cada serie del Año 1955 con su correspondiente del Año 1957

Tabla D

CONCEPTOS	AÑO 1955				AÑO 1957			
	Obre- ros ocu- pados	Horas- Obrero Traba- jados	Poder adqui- sitivo del monto de los salarios pa- gados	Volu- men físico de la produc- ción in- dustrial	Obre- ros ocu- pados	Horas- Obrero Traba- jados	Poder adqui- sitivo del monto de los salarios pa- gados	Volu- men físico de la produc- ción in- dustrial
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Ladrillos de máquina	4,85	4,70	3,59	2,86	5,08	5,09	3,97	3,30
Mosaicos	3,59	3,51	2,64	1,98	3,40	3,38	2,65	1,92
Fábricas de vidrio	4,29	4,31	3,08	2,52	4,30	4,46	3,44	2,80
m) METALES (EXCLUIDA MAQUINARIA)								
Artículos diversos de hierro .	5,02	4,41	3,67	3,12	5,30	4,94	4,42	3,22
Hojalaterías	5,02	4,86	3,82	3,45	5,31	5,10	4,40	3,53
Bulones, tornillos, etc. . . .	3,82	3,09	2,95	2,31	3,52	3,53	3,08	2,37
Caños de hierro y acero	6,47	6,33	5,11	4,10	6,59	6,86	5,96	5,86
Fundición y elaboración de cobre y bronce	5,21	4,74	4,01	2,91	5,04	4,83	4,30	3,81
Cocinas y otros artef. excluidos los electr.	5,41	5,18	4,04	3,24	5,42	5,28	4,54	3,90
Fund. y elabor. de hierro, ace- ro y otros metales	4,87	4,59	3,71	3,33	4,70	4,52	4,01	3,16
Hierro, galvanización	7,43	6,86	5,52	6,75	5,86	5,39	4,81	4,10
Fund. y elabor. de plomo, estaño y otros metales no ferrosos	5,23	5,04	3,81	4,23	5,60	5,62	4,49	5,05
Fundición de minerales de plomo, estaño y otros me- tales no ferrosos	4,96	4,83	3,89	1,97	4,86	4,82	4,66	2,52
Carpintería metálica.	4,35	3,89	3,08	1,60	4,02	3,71	3,05	1,71
n) VEHICULOS Y MAQUINARIA (EX- CLUIDA LA ELECTRIC.)								
Astilleros y talleres navales .	4,96	4,26	3,62	2,21	2,90	2,20	1,91	1,11
Armado de automotores y fabric. de carrocerías	4,99	4,60	3,69	6,60	5,46	5,27	4,58	10,44
Maquinaria y motores, exclu- idos los eléctricos	4,98	4,55	3,69	2,35	5,05	4,66	3,84	2,38
ñ) MAQUINARIA Y APARATOS ELEC- TRICOS								
Acumuladores, pilas y baterías	4,22	3,77	3,10	1,87	4,52	4,24	3,78	2,78
Artículos y aparatos divers. para electricidad	6,16	5,68	4,85	3,03	5,81	5,59	5,11	3,36
Lámparas y tubos para elec- tricidad	6,14	5,34	4,78	2,76	6,25	6,11	6,35	4,10
Motores eléctricos	5,23	4,73	4,06	3,19	4,94	4,28	4,30	1,75
Radiofonía	7,00	6,37	5,46	3,55	6,48	6,66	5,55	3,91
o) VARIOS								
Botones, boquillas, peines, etc.	4,31	3,98	3,24	2,17	3,73	3,49	3,01	1,72
Hielo	4,52	4,10	3,40	2,12	4,51	4,15	3,39	1,85
ELECTRICIDAD Y GAS								
Fábricas de electricidad . . .	5,26	4,71	4,41	2,77	5,74	5,21	5,37	2,98
Gas manufacturado	5,01	4,33	3,58	2,56	5,56	5,04	4,15	2,74

VITTORIO CASTELLANO

Sull'insieme delle distribuzioni doppie e triple risultanti dall'associazione una a una delle unità di due o tre distribuzioni semplici aventi lo stesso numero di unità (*)

SOMMARIO

- § 1 — GENERALITÀ I.1. Scopo del lavoro. I.2. Alcune precisazioni sul significato del termine « variabile » in matematica e in statistica. Variabile statistica e distribuzione statistica.
- § 2 — VARIABILI DOPPIE. 2.1. La classe delle variabili doppie definita dall'associazione di due variabili semplici. Le tre distribuzioni caratteristiche della classe. 2.2. Definizione della operazione di « trasporto » che trasforma una distribuzione doppia in un'altra dello stesso insieme. Esistenza di due sole distribuzioni lineari. 2.3. Caso delle variabili continue. Estensione al caso delle variabili continue empiriche. Classe di Fréchet delle funzioni di ripartizione doppie. 2.4. Funzioni di ripartizione massima e minima di una classe di Fréchet. 2.5. Estremo superiore e inferiore dell'insieme delle variabili doppie definite da due variabili semplici. Distanza di una distribuzione dell'insieme da una distribuzione estremante. 2.6. Valori estremi dei coefficienti di regressione e di correlazione, dei rapporti di correlazione, degli indici di connessione totale e globale di una classe di Fréchet di distribuzioni doppie. 2.7. Applicazione alla programmazione lineare e condizione sufficiente perchè le distribuzioni estremanti costituiscano la soluzione di un problema di programmazione lineare.
- § 3 — VARIABILI TRIPLE. 3.1. Distribuzioni triple risultanti dall'associazione di tre variabili semplici. Classe di Fréchet di una variabile tripla. Limitazione fondamentale per la funzione di ripartizione di una distribuzione della classe. 3.2. Rappresentazione continua delle

(*) La impostazione di questo lavoro ed i principali risultati relativi alle distribuzioni doppie sono stati esposti nella comunicazione « *Sulla variabile statistica doppia risultante dall'associazione di due date variabili statistiche semplici* », fatta al Seminario di Statistica metodologica dell'Istituto di Statistica di questa Facoltà, nella riunione del 26 gennaio 1957.

variabili triple empiriche. Le quattro distribuzioni lineari della classe di Fréchet di variabili triple. 3.3. Uniformizzazione della rappresentazione. Non esistenza di una distribuzione tripla la cui funzione di ripartizione coincida con l'estremo inferiore delle funzioni di ripartizione di una classe di Fréchet di distribuzioni triple. 3.4. Nuova definizione delle distribuzioni estremanti di una classe di Fréchet. Determinazione dell'insieme delle distribuzioni minime di una classe di Fréchet di distribuzioni triple uniformizzate nel caso $N = 10$. 3.5. Conclusione.

1. Il Fréchet, si è occupato delle variabili doppie che possono risultare dall'associazione di due variabili semplici, e R. Féron ha esteso le sue considerazioni al caso della associazione di tre variabili semplici (1). La novità della trattazione consiste, secondo noi, nell'uso delle funzioni di ripartizione, o curva cumulata delle frequenze, che dà luogo a risultati particolarmente espressivi, ma la tecnica dell'argomentazione è un così chiaro esempio di quelle deformazioni — direi professionali — che il matematico induce nella teoria della statistica, che non è pura esercitazione la rielaborazione dello stesso soggetto da parte dello statistico, per il quale la teoria è lo studio degli aspetti più generali dei fenomeni di massa che costituiscono il campo specifico del suo interesse (2).

2. Premettiamo qualche precisazione sull'impiego di alcuni termini cui la statistica e la matematica danno significati diversi.

Quantità variabile è una grandezza che assume più valori distinti. Tali grandezze sono dette in matematica variabili senza altro appellativo, e sono definite dall'insieme dei valori distinti (numeri reali) che esse assumono. Tale insieme può essere sempre posto in corrispondenza biunivoca con i punti di una linea (in particolare di una retta) ed è quindi sempre rappresentabile

(1) M. FRECHET — *Sur les tableaux de corrélation dont les marges sont données*, Annales Université Lyon, Section A, p. 53-77, 1951 — *Comptes rendus de l'Académie des Sciences*, t. 242, 1956; R. FERON: idem, *Cas de l'espace à trois dimensions*, in Publications de l'Institut de Statistique de l'Université de Paris, vol. V, fasc. I, 1956. Vedi anche A. RIZZI: *Osservazioni sulle classi di Fréchet delle funzioni di ripartizione a più variabili*.

(2) V. CASTELLANO — *Statistica e Matematica*, discorso inaugurale della XVIII Riunione della Società Italiana di Statistica, in «Atti della XVIII Riunione della S.I.S.», Roma, 1958.

con un certo insieme di punti di una retta. Pertanto una variabile matematica è un insieme lineare di punti, ed è, con tale insieme, discreta o continua, limitata o illimitata, e così via.

Variabile statistica è l'insieme potenziale dei valori che una grandezza empirica assume in un certo campo di osservazione nel quale la variabile è definita. I valori distinti che la grandezza può presentare, a priori, in un qualunque campo di osservazione, costituiscono le modalità della grandezza (o del carattere) e si identificano con la variabile matematica definita dalla grandezza. I valori effettivi, non necessariamente tutti distinti, che la grandezza presenta nell'insieme delle unità statistiche componenti un particolare campo di osservazione, coincidono con tutte o parte delle modalità del carattere, e definiscono una « distribuzione statistica ». Una distribuzione statistica può essere data sotto la forma di « distribuzione di frequenza »

$$f_i = f(x_i) \quad [1]$$

nella quale ciascuna frequenza f_i indica quante volte, nell'insieme dato, la grandezza X presenta la modalità x_i , ove $i = 1, 2$ p. u.

Altre forme sono:

quella della curva di graduazione:

$$x_i = x(i) \quad [2]$$

nella quale i rappresenta il posto che nella graduatoria crescente del carattere ha il valore x_i , convenuto che quantità uguali occupano posti successivi; e della distribuzione cumulata delle frequenze, o funzione di ripartizione, che ha la stessa rappresentazione grafica della [2], salvo lo scambio degli assi

$$i = i(x) \quad [3]$$

La variabile statistica è sempre un insieme finito e certamente discreto e limitato, anche se la variabile matematica corrispondente è continua. In quest'ultimo caso, i valori x_i essendo generalmente delle misure, essi sono affetti da un errore di approssimazione ε_i non superiore alla metà della distanza, costante, tra due modalità successive del carattere. Le modalità effettive del carattere sono costituite di solito da intervalli aventi per estremi i punti equidistanti $x_i \pm d$.

Le f_i quantità x_i non sono in realtà tutte uguali. Indicando con x_{ij} i valori esatti, occorrerebbe porre (3)

$$f_i = \sum_j f_{ij} \quad [4]$$

La [4] può anche, senza inconvenienti essere posta sotto la forma:

$$f_i = \int_{x_i-d}^{x_i+d} f(x) dx \quad [4']$$

ove non si dà alla [4'] altro significato che quello legittimo di frequenza delle quantità appartenenti all'intervallo $x_i \pm d$.

Le curve di distribuzione delle variabili statistiche sono definite per punti. È quasi sempre comodo dare continuità a tali rappresentazioni, sia dal punto di vista analitico che da quello della rappresentazione grafica. Ciò si fa in base a convenzioni quasi ovvie e a semplici ipotesi, che sono tuttavia da tener presenti quando si operino trasformazioni su tali rappresentazioni continue. (4)

In conclusione, caratteristica essenziale della variabile statistica è quella di essere sì discreta, ma di definire una funzione di densità. Essa deve essere quindi considerata un ente a due dimensioni.

3. L'associazione una a una delle N quantità di due variabili semplici X e Y aventi uno stesso numero di quantità, dà luogo a una variabile doppia $Z \equiv (X Y)$.

Se le quantità X_i fossero tutte diverse e similmente le Y_i le associazioni diverse possibili sarebbero $N!$ Queste si ottengono associando ad una disposizione di uno dei due caratteri tutte le permutazioni dell'altro. Tale numero si riduce se le quantità non sono tutte diverse tra loro e si riduce a 1 nel solo caso che tutte le quantità di un carattere siano uguali fra loro, cioè se la X o la Y si riducono a una costante. Escluso questo caso, sono certamente distinte le variabili doppie che si ottengono associando, nell'ordine, ai termini in graduatoria dell'una i termini dell'altra nello stesso ordine o nell'ordine inverso.

(3) V. CASTELLANO: *Sulle variabili divise in intervalli e le correzioni della media aritmetica e della varianza* in « Statistica » Anno XVI, n. 2, Aprile-Giugno 1956.

(4) V. CASTELLANO: *Sulla rappresentazioni grafica delle curve di distribuzione dei caratteri continui* in « Atti della XI Riunione della Soc. Ital. di Statistica, Roma 1951.

Una distribuzione doppia è descritta perfettamente dalla cosiddetta « tabella di correlazione » che meglio dovrebbe essere denominata tabella di associazione. Essa è una tavola a doppia entrata nella colonna e nella riga madri della quale sono le modalità X_i e Y_j dei due caratteri associati, nella colonna e riga estreme le frequenze N_{i0} e N_{0j} , tutte > 0 , corrispondenti e nelle caselle interne le frequenze n_{ij} delle varie coppie $X_i Y_j$.

Tutte le variabili doppie che risultano dall'associazione delle stesse variabili semplici costituiscono un insieme di tabelle aventi le stesse distribuzioni marginali (X_i, N_{i0}) , (Y_j, N_{0j}) . Se almeno una delle due variabili non è costante, (nel qual caso la tabella si riduce a una sola riga o colonna), sono certamente distinte le tabelle di graduazione e di contrograduazione. La prima di queste si costruisce ponendo nella 1^a casella in alto a sinistra

$$n_{11} = N_{10} \text{ oppure } n_{11} = N_{01}$$

a seconda che $N_{10} \leq N_{01}$.

Se $N_{10} = N_{01}$ in tutte le altre caselle della 1^a riga e della 1^a colonna le n_{ij} saranno nulle e si porrà :

$$n_{22} = \min (N_{20}, N_{02}) ;$$

se $N_{10} \leq N_{01}$, per esempio $N_{10} < N_{01}$ si porrà $n_{11} = N_{10}$ e $n_{1j} = 0$ (per $j = 2, 3, \dots$). Essendo così determinante tutte le frequenze

	x_1	x_2	...	x_i	...	x_r	
y_1	n_{11}	n_{21}	...	n_{i1}	...	n_{r1}	N_{01}
y_2	n_{12}	n_{22}	...	n_{i2}	...	n_{r2}	N_{02}
\vdots	
y_j	n_{1j}	n_{2j}	...	n_{ij}	...	n_{rj}	N_{0j}
\vdots	
y_s	n_{1s}	n_{2s}	...	n_{is}	...	n_{rs}	N_{rs}
	N_{10}	N_{20}	...	N_{i0}	...	N_{r0}	N

della prima colonna si applica lo stesso procedimento alla tabella che si ottiene dalla precedente sopprimendo in essa la prima colonna e le N_{10} unità in essa classificate.

Tale tabella risulta :

	x_2 x_3 ... x_i ... x_r	
y_1	n_{21} n_{31} ... n_{i1} ... n_{r1}	$N_{01} - N_{10}$
y_2	n_{22} n_{32} ... n_{i2} ... n_{r2}	N_{02}
\vdots	
y_j	n_{2j} n_{3j} ... n_{ij} ... n_{rj}	N_{0j}
\vdots	
y_s	n_{2s} n_{3s} ... n_{is} ... n_{rs}	N_{0s}
	N_{20} N_{30} ... N_{i0} ... N_{0s}	$N - N_{10}$

e in questa si porrà: $n_{21} = \min(N_{20}, N_{01} - N_{10})$ e $n_{2i} = 0$ per $j = 2, 3, \dots s$ se $N_{20} < N_{01} - N_{10}$ oppure $n_{i1} = 0$ per $i = 3, 4 \dots r$, se $N_{20} > N_{01} - N_{10}$.

Determinate così tutte le frequenze della prima riga o della prima colonna si passerà a determinare quelle della tabella che si ottiene sopprimendo nella tab. 2^a tale riga o colonna e così via.

Per tutte le coppie di valori i, j per le quali, posto

$$N_{i0} = \sum_j n_{ij}, N_{0j} = \sum_i n_{ij}$$

accada che $N_{i0} = N_{0j}$ si passa dalla tabella precedente a quella con $r - i$ colonne e $s - j$ righe sopprimendo insieme una riga e una colonna della tabella precedente.

Il procedimento ha termine quando si giunge ad una tabella con una sola riga o colonna, le cui frequenze di porranno uguali a quelle della riga o colonna marginale.

Se si parte dalla 1^a casella in basso a destra ponendo $n_{rs} = \min(N_{r0}, N_{0s})$ si arriva alla stessa tabella di associazione, perchè la tabella di cograduazione è unica.

Se si applica lo stesso procedimento a partire dalla prima casella in basso a sinistra o in alto a destra si arriva alla tabella di contrograduazione (1).

Segue dal procedimento usato che la tabella di cograduazione (e similmente quella di contrograduazione) consta di un insieme lineare di caselle contenenti frequenze $\neq 0$. Se, nel procedimento sopra illustrato, tutte le successive tabelle si ottengono dalle precedenti mediante soppressione di una riga oppure di una colonna, e il procedimento ha termine dopo la soppressione di h righe e di k colonne, ove si indichi con t il numero delle caselle delle quali consta la riga o la colonna della tabella residua, sarà

$$h = r - 1, t = s - k,$$

oppure

$$k = s - 1, t = r - h,$$

e comunque

$$h + k + t = r + s - 1 \quad [5]$$

(1) Il FRÉCHET, op. cit., dimostra laboriosamente che, date le due distribuzioni marginali, si può sempre costruire qualche tabella di correlazione. La esistenza di queste segue immediatamente dalla possibilità di associare una a una le qualità di due gruppi ugualmente numerosi di quantità e non aveva bisogno di essere provata anche prima che il Salvemini costruisse la tabella di cograduazione e di contrograduazione (T. SALVEMINI: *Sugli indici di omofilia*, in « Atti della Soc. It. di Statistica, 1939). Ma ciò che ci lascia veramente perplessi nella trattazione del Fréchet non è il fatto che il matematico, pur partendo da uno strumento strettamente statistico quale la tabella di correlazione abbia dimenticato che essa non è altro che la classificazione di coppie ordinate di quantità, ma il fatto che egli, considerando separatamente il caso delle distribuzioni discrete e delle distribuzioni continue, abbia attribuito il primo alla statistica e il secondo al calcolo delle probabilità. Ora le variabili casuali, oggetto precipuo del c.d.p. non sono che particolari variabili statistiche, collegate ai risultati possibili di una scelta a caso. E come le variabili statistiche, sono certamente discrete le variabili casuali empiriche, e solo per comodità si considerano generalmente continue le v.c. teoriche.

E infine, l'operazione di scelta, che caratterizza le variabili casuali nella classe più vasta delle variabili senza qualificazione, non ha nessun interesse nella trattazione. Perché allora tale qualificazione?

Il rilievo non è fatto certamente per il Fréchet, la cui statura scientifica consente licenze anche maggiori, ma per gli statistici in particolare, che si rivolgono al pubblico di coloro che applicano il metodo statistico, e in generale, a tutti coloro per i quali ciò che per qualcuno è « licenza » torna ad essere solamente un errore.

$h + k + t$ rappresenta il numero v delle caselle della tabella di cograduazione contenente frequenze non nulle, nella detta ipotesi di $N_{i0} \neq N_{0j}$ per ogni coppia i, j .

Tale numero si riduce di una unità ogni qualvolta nel passaggio da una tabella alla successiva, essendo $N_{i0} = N_{j0}$ si sopprimono insieme una linea e una colonna, e ciò non può avvenire più di $r - 1$ o $s - 1$ volte a seconda che $r \leq s$. Pertanto il numero v delle frequenze non nulle nella tabella di cograduazione è

$$\max(r, s) \leq v \leq r + s - 1 \quad [6]$$

4. Altra tabella di associazione che si usa considerare nella statistica teorica è la cosiddetta tabella di indipendenza nella quale la frequenza della coppia (X_i, Y_j) è definita dalla relazione

$$n'_{ij} = \frac{n_{i0} n_{0j}}{N} \quad [7]$$

ove

$$N = \sum N_{i0} = \sum N_{0j}.$$

Poichè d'ordinario il prodotto $n_{0i} n_{0j}$ non è multiplo intero di N non esisterà una v.s. doppia la cui distribuzione sia conforme alla [7], ma solo che si avvicini a tale distribuzione.

Le due variabili doppie di cui abbiamo determinato la distribuzione delle frequenze hanno la singolare proprietà di avere frequenza nulla in tutte le caselle della tabella di associazione, escluse v caselle costituenti un insieme lineare.

Se materializziamo la v.d. in N palline distribuite nelle r.s. caselle della tabella, tutte le variabili doppie compatibili coi vincoli imposti si otterranno spostando, per coppie, unità di righe (colonne) differenti, ciascuna nella propria riga (colonna), in maniera che se una passa dalla colonna i esima alla l esima, l'altra passa dalla colonna l esima alla colonna i esima. Così resta inalterato il numero delle unità nelle varie righe e colonne.

Nella tabella di cograduazione i trasferimenti possono avvenire solo dagli estremi della diagonale principale della tabella minima contenente le due caselle dalle quali due unità si trasferiscono, verso gli estremi della diagonale secondaria; mentre nella tabella di contrograduazione è l'opposto: ciò segue dalla osserva-

zione che la linea di distribuzione delle unità è non decrescente nel primo caso e non crescente nel secondo.

Osserviamo anche che se ci proponiamo di distribuire le N unità della tabella lungo un insieme lineare di caselle, cioè tale che da un estremo all'altro dell'insieme delle caselle si possa solo andare mediante cammini che appartengano interamente all'insieme e che attraversino tutte le caselle dell'insieme, ci accorgiamo che solo le due distribuzioni considerate godono di tale proprietà. Infatti condizione necessaria per un tale ordinamento è che ciascuna casella confini con due caselle dell'insieme salve le caselle estreme che debbono avere un lato in comune con una sola casella dell'insieme, e se l'ordinamento si comincia a realizzare da uno dei vertici della tabella si ha la tabella di cograduazione o quella di contrograduazione; se si comincia da un'altra casella delle righe o colonne estreme, o si ricade nelle associazioni precedenti, o si costruisce un insieme spezzato. Possiamo quindi dire che: tra tutte le variabili doppie aventi le stesse distribuzioni marginali solo due sono ordinate linearmente (e sono quelle che si ottengono cograduando o contrograduando i valori delle variabili semplici componenti).

Se i due caratteri sono continui e N è sufficientemente grande perchè aumentando il numero delle classi non restino caselle vuote nelle due distribuzioni marginali, al crescere di r e s le unità delle due variabili doppie considerate si dispongono su un nastro sempre più sottile di caselle.

5. Al limite, se consideriamo continue le due distribuzioni marginali, di densità $\varphi(x)$ e $\psi(y)$ e $N = \infty$, posto

$$\Phi(x) = \int_{-\infty}^x \varphi(u) du, \quad \Psi(y) = \int_{-\infty}^y \psi(v) dv, \quad \Phi(\infty) = \Psi(\infty) = 1$$

la equazione

$$\Phi(x) = \Psi(y) \quad [8]$$

definisce sul piano cartesiano la linea lungo la quale è distribuita la v.d. nel caso di cograduazione delle v.s. componenti mentre la equazione

$$\Phi(x) = 1 - \Psi(y) \quad [9]$$

definisce la linea di distribuzione delle v.d. nel caso di contrograduazione.

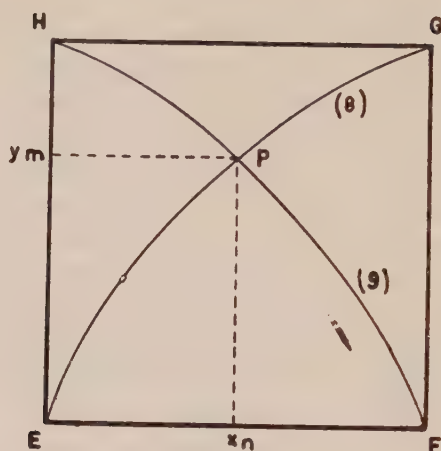


Fig. 1

Se $a \dashv \vdash b$ è il campo di variazione della v. X e $c \dashv \vdash d$ quello della v. Y , per cui:

$$\begin{array}{llll} \Phi(x) = 0 & \text{per } x \leq a & \Phi(x) = 1 & \text{per } x \geq b \\ \psi(y) = 0 & \text{per } y \leq c & \psi(y) = 1 & \text{per } y \geq d \end{array}$$

la variabile doppia $Z = (X, Y)$ è definita nel rettangolo $EFGH$ e la [8] rappresenta una curva monotona che va dal vertice E al vertice G , la [9] una curva monotona che va da H ad F . Le due curve si intersecano nel punto M nel quale

$$\Phi(x) = \psi(y) = 1 - \psi(y) = \frac{1}{2}$$

le cui coordinate sono le mediane delle due distribuzioni marginali.

In pratica, le variabili continue sono date per classi di valori. Se, in tale caso, supponiamo uniformemente distribuite in ciascuna classe le unità che essa contiene e suddividiamo ciascuna classe in altrettanti intervallini, ogni variabile semplice risulta trasformata in una successione d'intervallini di diverse ampiezze, in ciascuno dei quali cade una sola unità della distribuzione.

Se denominiamo « variabile continua empirica » ogni variabile continua così rappresentata, ci rendiamo facilmente conto che ogni variabile continua relativa a un numero finito N di unità è una « variabile continua empirica » e che ogni variabile empirica è continua quando si pensi data per classi di valori anzichè per valori singoli.

Le considerazioni fin qui fatte per l'associazione di variabili semplici continue valgono esattamente per le variabili continue empiriche, semplici, doppie, e così via, solo che si sostituisca al punto l'intervallo o l'elemento di volume dello spazio corrispondente alla dimensione della variabile.

Noi riferiamo pertanto il seguito delle nostre considerazioni alle variabili continue empiriche, senza rinunciare affatto alle generalità.

Consideriamo una qualsiasi delle v.d. aventi le date variabili componenti, e sia $f(x, y)$ la sua funzione di densità e $F(x, y)$ quella di ripartizione. Per comodità usiamo tali notazioni che ci permettono di trattare insieme il caso teorico di v. continua e quello generale delle v. discontinue. Gli adattamenti a questo secondo caso sono ovvii.

La $F(x, y)$ è definita dalla relazione

$$F(x, y) = \int_a^x \int_c^y f(u, v) du dv \quad [10]$$

e soddisfa alle seguenti condizioni

$$\begin{aligned} F(x, d) &= \Phi(x); & F(b, y) &= \psi(y) \\ F(a, c) &= F(a, y) = F(x, c) = 0; & F(b, d) &= 1 \end{aligned} \quad [11]$$

Indichiamo con F_1, F_2, F_3, F_4 , rispettivamente le frequenze delle coppie delle v.d. che cadono nei quattro rettangoli aventi un vertice in $P(x, y)$ interno al rettangolo $EF GH$ e vertici opposti rispettivamente in E, F, G e H .

Risulta immediatamente:

$$\begin{aligned} F_1 + F_4 &= \Phi(x); & F_1 &\leq \Phi(x) \\ F_1 + F_2 &= \psi(y); & F_1 &\leq \psi(y) \end{aligned}$$

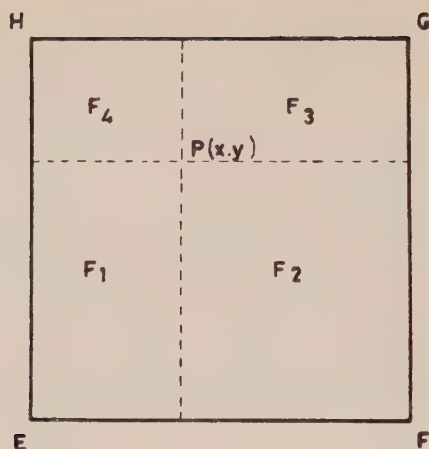


Fig. 2

$$F_2 + F_3 = 1 - \Phi(x)$$

$$F_1 - F_3 = \Phi(x) + \psi(y) - 1 \quad F_1 \geq \Phi(x) + \psi(y) - 1$$

e poichè $F_1 \equiv F(x, y)$, ed è, per definizione, $F_1 \geq 0$:

$$\max [0, \Phi(x) + \psi(y) - 1] \leq F(x, y) \leq \min [\Phi(x), \psi(y)] \quad [12]$$

Tutte le funzioni di ripartizione delle v.d. risultanti dall'associazione delle due date v. semplici soddisfano alla [12] e viceversa, ogni funzione $F(x, y)$ monotona rispetto a entrambe le variabili x e y , che soddisfi la [12] soddisfa anche la [11] ed è quindi funzione di ripartizione di una v.d. del detto insieme. La [12] costituisce pertanto la condizione necessaria e sufficiente affinché una funzione monotona di due variabili sia funzione di ripartizione di una v.d. dell'insieme.

Ogni funzione di ripartizione soddisfacente alla [12] definisce una distribuzione doppia associata alla data $\Phi(x)$ e $\psi(y)$, e viceversa. Denomineremo spesso in seguito per brevità « classe di Fréchet » l'insieme delle distribuzioni doppie definito dall'associazione di due variabili semplici (5).

(5) Non riteniamo però di poter accettare tale denominazione in generale e non la abbiamo quindi adottata per il titolo di questa nota, per un

6. Consideriamo le regioni nelle quali le curve (o l'insieme lineare di caselle, nel caso discontinuo) definite dalle equazioni [8] e [9], dividono il rettangolo $EFGH$.

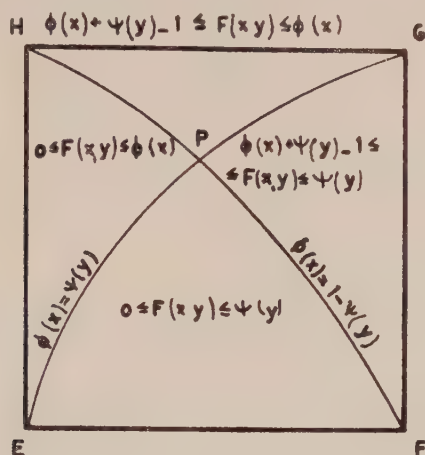


Fig. 3

La differenza $\Phi(x) - \psi(y)$ cambia segno sulla EPG : nella regione EPG è sempre $\Phi(x) > \psi(y)$ e nella EGH è $\Phi(x) < \psi(y)$, similmente sulla HPF cambia segno la differenza $\Phi(x) + \psi(y) - 1$ e poichè essa è funzione crescente di x e di y , è

$$\Phi(x) + \psi(y) - 1 < 0 \quad \text{in} \quad EFH$$

e

$$\Phi(x) + \psi(y) - 1 > 0 \quad \text{in} \quad HFG$$

nostro principio di coerenza scientifica ed una più generale esigenza di chiarezza a entrambi i quali non intendiamo rinunciare.

La scienza è prodotto eminentemente sociale e la individuazione di un teorema, una esperienza, una legge, mediante il nome di una persona contrasta col naturale processo di oggettivazione della scienza, ma in un ristretto ambiente accademico, nel quale le idee vengono vagliate e filtrate è dimostrazione di umani sentimenti di estimazione e di riconoscenza che sono anche altamente sociali se limitati ai casi di eccezionale significazione. In tutti i campi però, quando si verifichi inflazione, e in particolare in quello della statistica, nel quale più difficile è seguire i filoni delle idee, tale maniera di denominare diventa solo un contributo di più alla formazione e all'espandersi di quel gergo scientifico che è il prodotto più diretto

Tenuto conto di ciò la [12] può essere precisata come nella tabella seguente. La seconda e la terza colonna della tabella definiscono rispettivamente la funzione di ripartizione minima F_m

Regione	Estremo della funzione $F(x, y)$	
	inferiore	superiore
EPF	0	$\psi(y)$
FPG	$\Phi(x) + \psi(y) - 1$	$\psi(y)$
GPH	$\Phi(x) + \psi(y) - 1$	$\Phi(x)$
HPE	0	$\Phi(x)$

e la funzione di ripartizione massima F_M della classe di Fréchet definita dalle $\Phi(x)$ e $\psi(y)$, rappresentate nelle figg. 4 e 5.

della specializzazione tecnica e della industrializzazione delle scienze, nella quali le esigenze dell'economia del tempo irresistibilmente prendono anche da noi il sopravvento sull'economia dei concetti, come si rileva dallo scendere progressivo della inquadratura sistematica della produzione delle nostre nuove leve.

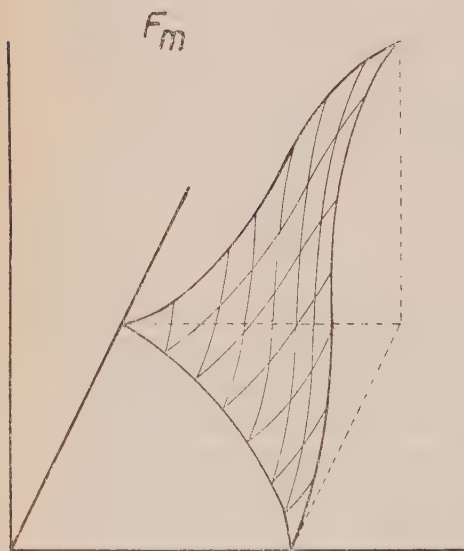


Fig. 4

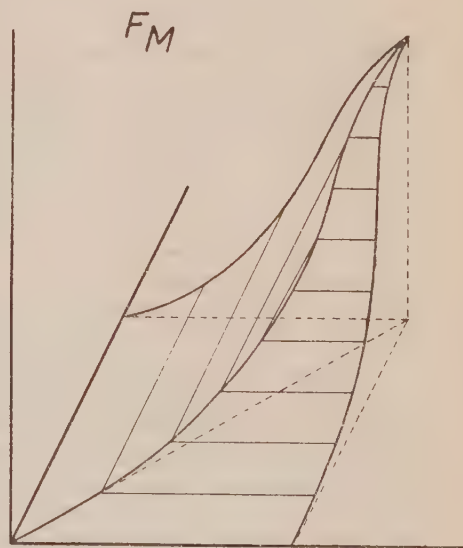


Fig. 5

La F_M è una superficie rigata lungo due segmenti di rette da ogni piano $F(x, y) = k$ ove $0 \leq k \leq 1$. Le proiezioni ortogonali di tali segmenti sul piano xy , hanno tutte un estremo sulla $\Phi(x) = \psi(y)$ [8]. Ciò vuol dire che la v.d. Z_m corrispondente alla F_M è tutta distribuita lungo la curva definita dalla [8].

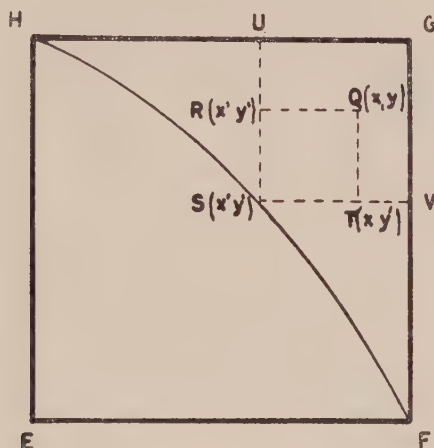


Fig. 6

La v.d. corrispondente alla F_m non ha unità nella regione EFH perchè in questa è $F_m = 0$. Dico che essa non ha unità nemmeno nella regione FGH , e quindi è distribuita solo lungo la curva definita dalla

$$\Phi(x) + \psi(y) - 1 = 0 \quad [9]$$

E infatti se la v. Z_m corrispondente avesse una unità in un punto $Q(x, y)$ qualunque di FGH , fuori della [9] sarebbe una variabile dell'insieme anche la Z' che si ottiene portando detta unità da $Q(x, y)$ in $R(x', y)$ con $x' < x$, ugualmente situato nella FGH , e $S(x', y')$ situato sulla [9] in $T(x, y')$, dato che la nuova variabile Z' soddisfa alle stesse condizioni ai margini. Nel campo $QRST$ contorno RQT escluso è $F' < F_m$ se F' è la funzione di ripartizione della Z' , mentre nel residuo campo $EFGH - QRST$ è $F' = F_m$.

La F' sarebbe quindi minore della F_m e ciò è assurdo perchè la F_m è la minima funzione della classe di Fréchet.

Se, in base alla definizione di funzione di ripartizione di una variabile doppia chiamiamo tabella di ripartizione la tabella che si ricava da quella di frequenza scrivendo in ciascuna casella (ij) la somma N_{ij} della frequenza relativa a tutte le caselle che hanno indici non superiori agli indici della data, talchè :

Tabella di ripartizione

	1	2	3	...	i	...	r
1	N_{11}	N_{21}	N_{i1}	...	N_{r1}	
2	N_{12}	N_{22}	N_{i2}	...	N_{r2}	
3	N_{13}	N_{23}	N_{i3}	...	N_{r3}	
\vdots	\vdots					
j	N_{1j}	N_{2j}	...	N_{ij}	...	N_{rj}	
\vdots	\vdots					
s	N_{1s}	N_{2s}	...	N_{is}	...	N_{rs}	

$$N_{ij} = \sum_{u=1}^i \sum_{v=1}^j n_{uv}$$

da cui

$$N_{i1} = \sum_{u=1}^i n_{u1} \quad N_{1j} = \sum_{v=1}^j n_{1v}$$

$$N_{rj} = \sum_{v=1}^j n_{rv} \quad N_{is} = \sum_{v=1}^i n_{sv}$$

$$N_{rs} = N$$

possiamo enunciare il seguente

Teor: Tra la tabella di frequenza delle v.d. risultanti dalla associazione delle stesse v.s. la tabella di cograduazione corrisponde alla tabella di ripartizione massima e la tabella di contrograduazione alla tabella di ripartizione minima.

7. Conveniamo di designare col nome di estremo superiore e inferiore dell'insieme di Fréchet le variabili doppie che si ottengono mediante cograduazione e contrograduazione delle due va-

riabili semplici che definiscono l'insieme. Tali estremi costituiscono, come abbiamo visto, le due soluzioni, « lineari », del problema dell'associazione di due variabili semplici e possono considerarsi pertanto anche come soluzioni degeneri in quanto le due distribuzioni doppie di frequenze si riducono a due insiemi lineari e quindi entrambi tali estremanti dell'insieme sono in effetti variabili semplici. Se, ad esempio, le due distribuzioni marginali sono identiche, per il che, posta l'origine in E e assunte come unità di misura sui due assi $\overline{EF} = 1$ e $\overline{EH} = 1$, risulta $\varphi(x) = \psi(y)$, per ogni $x = y$; la v. estremante superiore dell'insieme di Fréchet risulta definita lungo lo diagonale del quadrato $EF GH$ e la sua funzione di densità è $f = \frac{1}{\sqrt{2}} \varphi(x)$. Se la φ è simmetrica la stessa funzione di densità ha anche la estremante minima, sulla diagonale secondaria.

In particolare, se le due distribuzioni marginali sono normali, normale è anche la distribuzione delle due funzioni estremanti, e il suo scarto quadratico medio è quello delle funzioni marginali moltiplicato per $\sqrt{2}$.

Abbiamo già detto che il passaggio da una ad altra distribuzione dell'insieme si effettua scambiando di colonna o di riga compatibilmente coi vincoli, almeno due unità appartenenti a righe e a colonne diverse. Se assumiamo come distribuzione base la estremante inferiore, per ogni unità di una riga superiore che si sposta verso destra una o più unità di una riga inferiore dovranno spostarsi verso sinistra, di tante caselle, complessivamente, di quante caselle si è spostata la prima. Analogamente, per ogni unità che si sposti verso il basso in una colonna di sinistra, una unità di una colonna di destra dovrà sposarsi verso l'alto, e di tante caselle, complessivamente, di quante si è spostata la prima unità.

Se consideriamo unitaria la distanza tra due caselle successive, sia di una stessa riga che di una stessa colonna, segue che le unità di ciascuna distribuzione dell'insieme non sono necessariamente nello stesso numero in ciascuna delle due regioni nelle quali le

$$\Phi(x) = \psi(y) \quad [8]$$

$$0 \quad \Phi(x) + \psi(y) = 1 \quad [9]$$

dividono il campo di definizione dell'insieme, ma esse sono distribuite in maniera tale che la somma delle distanze delle unità dalla [8] e dalla [9] secondo le righe o le colonne, è costantemente nulla.

Se indichiamo rispettivamente con le lettere α e β le curve definite dalla [8] e [9] sul piano coordinato $x y$, talchè $(x_\alpha y_\alpha)$ siano le coordinate del punto generico della [8] e $(x_\beta y_\beta)$ quelle del punto generico della [9], tutte le distribuzioni dell'insieme soddisfano alle equazioni:

$$D_x = \int_c^d \int_{x_\alpha}^b (x - x_\alpha) f(x, y) dx dy =$$
[13]

$$= \int_c^d \int_a^{x_\alpha} (x_\alpha - x) f(x, y) dx dy$$

$$D_y = \int_a^b \int_{y_\alpha}^d (y - y_\alpha) f(x, y) dy dx =$$
[14]

$$= \int_a^b \int_c^{y_\alpha} (y_\alpha - y) f(x, y) dy dx$$

e le due analoghe che si ottengono sostituendo β al posto di α .
Se poniamo:

$$D = D_x + D_y$$
[15]

a ciascuna distribuzione dell'insieme viene associato un numero D che esprime la sua distanza dalla linearità.

Assunta come distribuzione base la distribuzione estremamente minima risulta:

$$D_\beta = 0 \quad \text{e} \quad D_\alpha = \max D$$
[16]

Ad ogni valore di D corrispondono in generale più distribuzioni.

8. In una classe di Fréchet di distribuzioni doppie, i coefficienti di regressione, il coefficiente di correlazione, il rapporto di correlazione, gli indici di connessione totale e globale e in generale ogni quantità caratteristica di una distribuzione doppia assume un insieme di valori che denomineremo, per semplicità, insieme di Fréchet della data caratteristica.

Dal fatto che le due distribuzioni estremanti di una classe di Fréchet sono co-o contro-graduate e lineari seguono le seguenti facili conseguenze :

a) per la distribuzione estremante inferiore β , la covarianza

$$p = \frac{\sum (x - \bar{x})(y - \bar{y})}{n} \quad [17]$$

assume il valore minimo della classe :

$$p_{\beta} = \min p \quad [18]$$

e per la distribuzione estremante superiore assume il valore massimo :

$$p_{\alpha} = \max p \quad (6) \quad [19]$$

b) per entrambe tali distribuzioni estremanti i rapporti di correlazione :

$$\rho_{y/x} = \sqrt{1 - \frac{\sum (y_{j/\lambda} - \bar{y}_j)^2}{\sum (y_{j/\lambda} - \bar{y})^2}} \quad \text{e} \quad \rho_{x/y} = \sqrt{1 - \frac{\sum (x_{i/\lambda} - \bar{x}_i)^2}{\sum (x_{i/\lambda} - \bar{x})^2}} \quad [20]$$

sono entrambi uguali al loro massimo assoluto 1 :

$$\rho_{\alpha} = \rho_{\beta} = \max \rho = 1 \quad [21]$$

(6) Infatti per un teorema fondamentale sulle medie ponderate (v. CASTELLANO : *Sulla teoria formale di alcune medie ferme basali*, in Atti della IX, X, XI Riunione della Soc. Italiana di statistica, Roma, 1950) in tali due casi è rispettivamente minima o massima la quantità

$$\frac{M(x - \bar{x})y}{M(y)} = p/M(y) \quad \text{mentre} \quad M(y) = \text{cost.}$$

Meno semplice è la dimostrazione che di questa stessa proprietà dà il FRÉCHET (*Les tableaux de corrélation et les programmes linéaires* in *Revue de l'Institut International de Statistique*, vol. 25, n. 1-3 La Haye 1957).

nel caso teorico di distribuzione continua e per le distribuzioni empiriche di caratteri continui quando la divisione delle variabili in intervalli è tale che due caselle delle distribuzioni estremanti α e β non abbiano mai un lato in comune.

c) Gli indici quadratici totale e globale di connessione (7), gli indici quadratici di connessione totale delle intensità medie di un carattere alle modalità dell'altro:

$${}^2C_{my/x} = \sqrt{\frac{\sum (\bar{y}_j - \bar{y})^2}{\sum (y_{j\lambda} - \bar{y})^2}} \quad {}^2C_{mx/y} = \sqrt{\frac{\sum (\bar{x}_i - \bar{x})^2}{\sum (x_{i\lambda} - \bar{x})^2}} \quad [22]$$

e gli indici quadratici di connessione globale:

$${}^2C_{mgy/x} = \sqrt{\frac{\sum (\bar{y}_j - \bar{y}_\lambda)^2}{\sum (y_{j\mu} - y_{\lambda\mu})^2}} \quad {}^2C_{mgx/y} = \sqrt{\frac{\sum (\bar{x}_i - \bar{x}_l)^2}{\sum (x_{iv} - x_{lv})^2}} \quad [23]$$

che si ottengono dai precedenti sostituendo agli scostamenti quadratici medi le differenze medie, assumono anch'essi il loro valore massimo 1:

$${}^2C_{m\alpha} = {}^2C_{m\alpha} = {}^2C_{m\beta} = {}^2C_{m\beta} = 1 \quad [24]$$

sotto le stesse condizioni espresse al punto b).

Pertanto:

gli estremi della classe del Fréchet del coefficiente di regressione, coefficiente di correlazione, indice quadratico di omofilia, rapporto di correlazione, indice quadratico di connessione, sono i valori di tali indici relativi alle distribuzioni estremanti della classe di Fréchet delle relative distribuzioni, le quali vengono ordinate da tutti questi indici in base alla « distanza » dalla distribuzione di « indipendenza » per la quale tutti questi indici assumono il valore zero.

In tale accezione è lecito dare il segno di p anche al rapporto di correlazione e agli indici di connessione, che assumono, come

(7) V. CASTELLANO: *Contributo alle teorie della correlazione e della connessione tra due variabili*, in « Metron », vol. XVIII, n. 3-4, 1957.

gli altri, il significato di indici non ordinali di cograduazione, cioè di indici di concordanza.

Si osservi esplicitamente che nessuno di tali indici di connessione assumerà, in questo nuovo significato, i valori estremi di ∓ 1 nei casi di contrograduazione e di cograduazione per le distribuzioni empiriche, necessariamente discontinue e di numerosità limitate, nei quali la corrispondenza tra le modalità dei due caratteri associati non può, in generale, essere biunivoca.

9. La osservazione che qualsiasi distribuzione dell'insieme si ottiene dalla estremante inferiore mediante trasporto di coppie di unità dagli estremi della diagonale secondaria di opportuni minori che hanno almeno un vertice sulla estremante inferiore agli estremi della diagonale principale; e dalla estremante superiore mediante il passaggio inverso, permette di individuare tali due distribuzioni anche come soluzioni d'interessanti problemi di massimo e minimo (8).

	2	4	6	4	7	2	5	
g		1		1			2	2
f		2		1	1	1	3	4
e	1		4		5	1	1	6
d	1		1	4	2	2	4	7
c		1	2		1			2
b		1	3	1	3			4
a	2	3	2	1	2			5
	1	2	3	4	5	6	7	

(8) vedi A. HERZEL: *Le tables de co- e contrograduazione e la programmazione lineare*, presentato e discusso nella Riunione del 27 aprile 1957 del Seminario di Statistica Metodologica dell'Istituto di Statistica, della Facoltà di Scienze Statistiche, Demografiche e Attuariali dell'Università di Roma, di prossima pubblicazione nella serie dell'Istituto di Statistica della Facoltà di Scienze Statistiche, Demografiche e Attuariali della Università di Roma; e M. FRÉCHET, *Le tableaux* ecc. già citato nella nota (6).

Il procedimento per ottenere dalla distribuzione indicata sopra quella estremante superiore è descritto nelle righe seguenti:

$$\begin{aligned} & 3a-1a, \quad 1d-3d; \quad 3a-2a, \quad 2c-3c; \quad 4a-1a, \quad 1e-4e; \\ & 2(5a-2a, \quad 2f-5f); \quad 4b-2b, \quad 2g-4g; \quad 3(5b-3b, \quad 3e-5e); \quad 5c-3c, \quad 3e-5e; \\ & 6d-4d, \quad 4e-6e; \quad 6d-4d, \quad 4f-6f; \quad 2(7d-4d, \quad 4g-7g); \\ & 2(7d-5d, \quad 5f-7f); \quad 7e-5e, \quad 5f-7f; \end{aligned}$$

Sia la variabile doppia λ_{ij} definita nello stesso campo di un insieme di Fréchet e si voglia determinare la distribuzione f_{ij} dell'insieme che rende massima o minima la quantità

$$S = \sum_{ij} \lambda_{ij} f_{ij} \quad [25]$$

Se nella serie di trasferimenti di coppie necessari per passare da una data distribuzione f_{ij} ad una estremante la S risulta monotona, allora tali distribuzioni estremanti costituiscono il minimo e il massimo cercato. Ciò avviene certamente se in tutti i minori della matrice λ_{ij} che hanno almeno una delle quattro caselle ai vertici su una distribuzione estremante, la somma degli estremi della diagonale principale è sempre non minore o non maggiore della somma degli estremi della diagonale secondaria. E precisamente se:

$$\lambda_{ij} + \lambda_{i+m,j+n} \geq \lambda_{i,j+n} + \lambda_{i+m,j} \quad [26]$$

la estremante superiore rappresenta un massimo per la S e la estremante inferiore un minimo, mentre se è

$$\lambda_{ij} + \lambda_{i+m,j+n} \leq \lambda_{i,j+n} + \lambda_{i+m,j} \quad [27]$$

la estremante superiore realizza il minimo di S e la estremante inferiore il massimo.

Nel campo continuo la funzione di densità della variabile doppia $\lambda(x, y)$ deve soddisfare alla stessa [26] e [27] e perciò è sufficiente che in tutti i punti del campo di definizione sia

$$\lambda(x + \Delta x, y + \Delta y) \begin{matrix} \geq \\ \leq \end{matrix} \lambda(x, y + \Delta y) + \lambda(x + \Delta x, y) - \lambda(x, y)$$

e quindi in definitiva condizione sufficiente perchè l'estremante superiore dell'insieme di Fréchet realizzi il massimo o il minimo di S è che

$$\lambda(x + \Delta x, y + \Delta y) - \lambda(x, y + \Delta y) + \lambda(x, y) - \lambda(x + \Delta x, y) \begin{matrix} \geq \\ \leq \end{matrix} 0$$

Cioè :

$$\lambda'_x(x, y + \Delta y) - \lambda'_x(x, y) \begin{matrix} \geq \\ \leq \end{matrix} 0$$

In definitiva :

$$\lambda''_{xy} \begin{matrix} \geq \\ \leq \end{matrix} 0 \quad [28]$$

10. Classe di Fréchet d'una variabile tripla.

Diremo similmente classe di Fréchet di una variabile tripla $\Lambda(X, Y, Z)$ l'insieme di tutte le distribuzioni triple definite dall'associazione di tre distribuzioni semplici X, Y, Z .

Siano $\Phi(x), \psi(y)$ e $\chi(z)$ le funzioni di ripartizione delle tre variabili semplici X, Y, Z .

Le variabili X, Y e Z definiscono, due a due, tre classi di Fréchet di variabili doppie contenute nell'insieme delle distribuzioni triple della classe. Tali classi sono due a due indipendenti nel senso che ogni distribuzione (X, Y) è associabile con qualunque distribuzione delle (X, Z) o delle (Y, Z) per definire una distribuzione della classe $\Lambda(X, Y, Z)$, la quale definisce la particolare distribuzione della terza classe compatibile con le distribuzioni date dalle due prime classi.

Se i campi di variazione delle tre variabili, X, Y, Z sono tutti e tre limitati inferiormente e superiormente, ogni unità di una delle distribuzioni triple della classe è rappresentabile con un punto di un parallelepipedo rettangolo con tre facce parallele ai piani coordinati, e tre sui piani coordinati, se assumiamo come origine del triedro di riferimento il punto avente come coordinate gli estremi inferiori delle tre variabili.

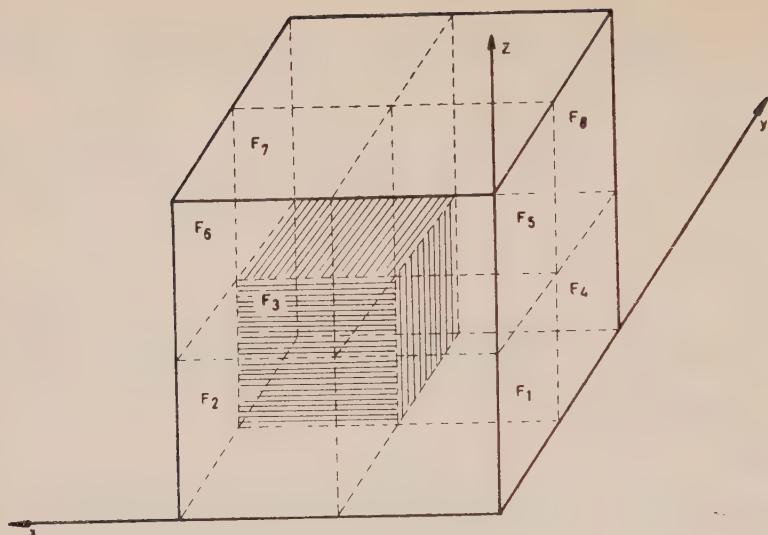


Fig. 7

I piani paralleli ai piani coordinati per un punto interno del parallelepipedo, dividono questo in 8 parallelepipedi che abbiamo numerato come in figura.

Se indichiamo con $F_i(X, Y, Z)$ la frequenza relativa delle unità della distribuzione tripla contenuta nel parallelepipedo i^{esimo} e similmente le frequenze relative delle unità delle distribuzioni doppie componenti che appartengono alle facce di tali parallelepipedi situate sui piani coordinati, si ha subito

$$F_1(X, Y, Z) = F_1(X, Y) - F_6(X, Y, Z)$$

$$F_1(X, Y, Z) = F_1(X, Z) - F_4(X, Y, Z)$$

È chiaro che $F_1(X, Y, Z)$ coincide colla funzione di ripartizione della variabile tripla.

D'altra parte è :

$$\max F_1(X, Y, Z) = \min(\Phi, \psi, \chi)$$

$$\max F_2(X, Y, Z) = \min(1 - \Phi, \psi, \chi)$$

$$\begin{aligned}
\max F_4 (X, Y, Z) &= \min (\Phi, 1 - \psi, \chi) \\
\max F_5 (X, Y, Z) &= \min (\Phi, \psi, 1 - \chi) \\
\max (0, \Phi + \psi - 1) &\leq F_1 (X, Y) \leq \min (\Phi, \psi) \\
\max (0, \Phi + \psi - 1) &\leq F_1 (X, Z) \leq \min (\Phi, \chi) \\
\max (0, \Phi + \psi - 1) &\leq F_1 (X, Z) \leq \min (\Phi, \chi) \\
\max (0, \psi + \chi - 1) &\leq F_1 (Y, Z) \leq \min (\Psi, \chi)
\end{aligned}
\tag{29}$$

e quindi segue che :

$$\begin{aligned}
\min F_1 (Y, Z, X) &\geq \min F_1 (Y, Z) = \max F_2 (X, Y, Z) = \\
&= \max (0, \psi + \chi - 1) - \min (1 - \Phi, \psi, \chi) \\
&\geq \min F_1 (X, Z) - \max F_4 (X, Y, Z) = \\
&= \max (0, \Phi - 1) - \min (\Phi, 1 - \psi, \chi) \\
&\geq \min F_1 (X, Y) - \max F_5 (X, Y, Z) = \\
&= \max (0, \Phi + \psi - 1) - \min (\Phi, \psi, 1 - \chi)
\end{aligned}
\tag{30}$$

I cilindri

$$\Phi + \psi - 1 = 0$$

$$\Phi + \chi - 1 = 0$$

$$\psi + \chi - 1 = 0$$

dividono lo spazio rappresentativo in 8 regioni.

Nella regione definita dalle :

$$\left\{ \begin{array}{l} \Phi + \psi - 1 < 0 \\ \Phi + \chi - 1 < 0 \\ \psi + \chi - 1 < 0 \end{array} \right.$$

è

$$\min F_1 (X, Y, Z) \geq 0$$

e quindi segue dalla prima delle [29]

$$0 \leq F_1 (X, Y, Z) \leq \min (\Phi, \psi, \chi).$$

Nella regione

$$\left\{ \begin{array}{l} \Phi + \psi - 1 < 0 \\ \Phi + \chi - 1 < 0 \\ \psi + \chi - 1 > 0 \end{array} \right.$$

nella quale è

$$\left\{ \begin{array}{l} \psi < 1 - \Phi \\ \Phi < 1 - \chi < \psi \\ \Phi < 1 - \psi < \chi \\ \chi < 1 - \Phi \end{array} \right.$$

e quindi

$$\Phi < \left\{ \begin{array}{l} 1 - \psi < \chi \\ 1 - \chi < \psi \end{array} \right\} < 1 - \psi \quad [31]$$

deve essere per le [29] e [30]

$$\min F_1(X, Y, Z) \geq \max(0, \psi - 1, \chi - 1)$$

e quindi per le [31]

$$0 \leq F_1(X, Y, Z) \leq \Phi$$

Analogamente si ottiene:

$$0 \leq F_1(X, Y, Z) \leq \psi$$

in

$$\left\{ \begin{array}{l} \Phi + \psi - 1 < 0 \\ \Phi + \chi - 1 > 0 \\ \psi + \chi - 1 < 0 \end{array} \right.$$

poichè

$$\psi < \left\{ \begin{array}{l} 1 - \Phi < \chi \\ 1 - \chi < \Phi \end{array} \right\} < 1 - \psi \quad [32]$$

in

$$\left\{ \begin{array}{l} \Phi + \psi - 1 > 0 \\ \Phi + \chi - 1 < 0 \\ \psi + \chi - 1 < 0 \end{array} \right.$$

poichè

$$\chi < \left\{ \begin{array}{l} 1 - \Phi < \psi \\ 1 - \psi < \Phi \end{array} \right\} < 1 - \chi \quad [33]$$

Nella regione :

$$\left\{ \begin{array}{l} \Phi + \psi - 1 < 0 \\ \Phi + \chi - 1 > 0 \\ \psi + \chi - 1 > 0 \end{array} \right.$$

nella quale è

$$\left\{ \begin{array}{l} \psi < 1 - \Phi < \chi \\ \Phi < 1 - \psi < \chi \\ 1 - \chi < \psi \\ 1 - \chi < \Phi \end{array} \right.$$

e quindi :

$$1 - \chi < \left\{ \begin{array}{l} \Phi < 1 - \psi \\ \psi < 1 - \Phi \end{array} \right\} < \chi \quad [34]$$

$$0 \leq F_1(X, Y, Z) \leq \min(\Phi, \psi)$$

e analogamente è :

$$0 \leq F_1(X, Y, Z) \leq \min(\Phi, \chi)$$

in

$$\left\{ \begin{array}{l} \Phi + \psi - 1 > 0 \\ \Phi + \chi - 1 < 0 \\ \psi + \chi - 1 > 0 \end{array} \right.$$

poichè è

$$1 - \psi < \left\{ \begin{array}{l} \Phi < 1 - \chi \\ \chi < 1 - \Phi \end{array} \right\} < \psi \quad [35]$$

$$0 \leq F_1(X, Y, Z) \leq \min(\Psi, \chi)$$

in

$$\begin{cases} \Phi + \psi - 1 > 0 \\ \Phi + \chi - 1 > 0 \\ \psi + \chi - 1 < 0 \end{cases}$$

poichè è

$$1 - \Phi < \left\{ \begin{array}{l} \psi < 1 - \chi \\ \chi < 1 - \psi \end{array} \right\} < \Phi \quad [36]$$

Infine nella regione :

$$\begin{cases} \Phi + \psi - 1 > 0 \\ \Phi + \chi - 1 > 0 \\ \psi + \chi - 1 > 0 \end{cases}$$

nella quale è

$$\begin{cases} 1 - \Phi < \psi \\ 1 - \Phi < \chi \\ 1 - \psi < \Phi \\ 1 - \psi < \chi \\ 1 - \chi < \Phi \\ 1 - \chi < \psi \end{cases}$$

risulta :

$$\max(0, \Phi + \psi + \chi - 2) \leq F_1(X, Y, Z) \leq \min(\Phi, \psi, \chi)$$

Se si osserva che :

$$\Phi + \psi + \chi - 2 = \begin{cases} \Phi - (1 - \psi) - (1 - \chi) \\ \psi - (1 - \chi) - (1 - \Phi) \\ \chi - (1 - \Phi) - (1 - \psi) \end{cases}$$

si rileva subito che

$$\Phi + \psi + \chi - 2 < 0$$

in tutte le regioni precedentemente considerate, salvo l'ultima, e quindi che vale in tutto lo spazio rappresentativo che

$$\max (0, \Phi + \psi + \chi - 2) \leq F(X, Y, Z) \leq \min (\Phi, \psi, \chi) \quad [37]$$

II. La interpretazione dei risultati precedenti risulta più chiara, senza perdere in generalità, se ci riferiamo ad una rappresentazione standardizzata della variabile tripla (X, Y, Z) .

Consideriamo le seguenti distribuzioni delle 3 variabili semplici

X		Y		Z	
x_i	φ_i	y_i	ψ_i	z_i	χ_i
1	1	1	4	1	2
2	6	2	2	2	4
3	3	3	1	3	2
		4	3	4	2

Assumiamo i valori delle tre variabili come centri di intervalli di ampiezza unitaria e come vertice di un nuovo triedro di riferimento avente gli assi paralleli ai precedenti, il punto

$$(0,5 - 0,5 - 0,5) .$$

Assumiamo come unità sui tre assi i seguenti

$$0,5 - 3,5, 0,5 - 4,5, 0,5 - 4,5$$

e rappresentiamo tali unità con lo stesso segmento sui tre assi. Le tre variabili diventano :

X		Y		Z	
x_i	φ_i	y_i	ψ_i	z_i	χ_i
0-1/3	1	0-1/4	4	0-1/4	2
1/3-2/3	6	1/4-2/4	2	1/4-2/4	4
2/3-3/3	3	2/4-3/4	1	2/4-3/4	2
		3/4-4/4	3	3/4-4/4	2

Se ora suddividiamo ciascun intervallo in tante parti uguali quante sono le unità che appartengono ad esso e in ciascun intervallino poniamo una delle unità della classe, ciascuna variabile risulta divisa in 10 intervalli (disuguali) e $\varphi_i = \psi_i = \chi_i = 1$ per qualunque i .

Tali variabili definiscono nel cubo unità, di vertici opposti (000)-(111), mille parallelepipedi rettangoli raggruppati in 10 strati di 100 secondo gli intervalli di ciascuna delle tre variabili semplici. Ogni distribuzione tripla dell'insieme ha una ed una sola unità in ciascuno dei 10 strati di tutti e tre gli ordinamenti.

Poichè ogni distribuzione tripla (X, Y, Z) si proietta secondo gli assi coordinati nelle tre variabili doppie (X, Y) , (Y, Z) , (X, Z) delle classi di Fréchet definite da due delle variabili semplici date

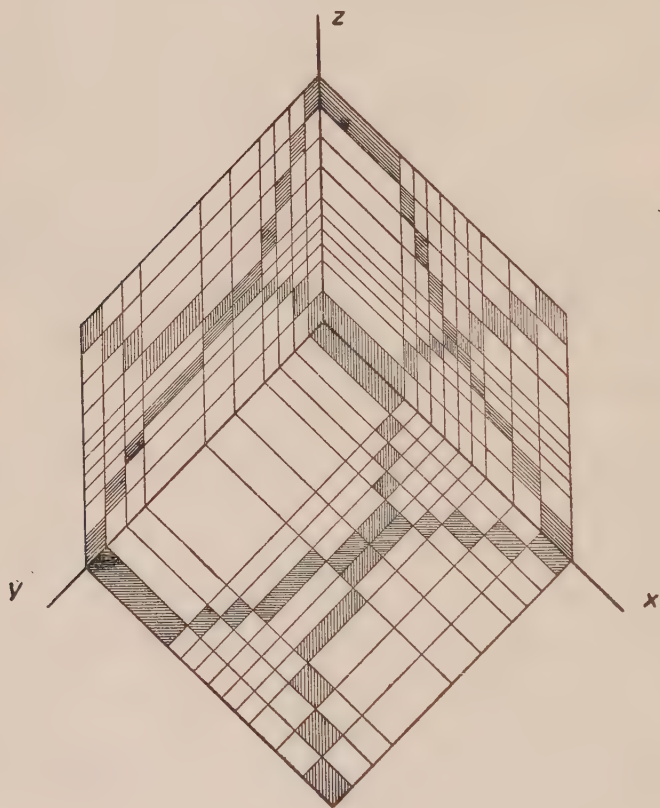


Fig. 8

e viceversa, e due sole di queste definiscono la variabile tripla, segue che la variabile tripla (X, Y, Z) ha quattro distribuzioni lineari, le quali hanno come componenti le distribuzioni estremanti delle tre variabili doppie componenti. Le funzioni di ripartizioni

di tali 4 variabili triple sono * definite in ogni punto dalle relazioni

$$\begin{aligned} F(\overrightarrow{X Y Z}) &= \min(\Phi, \psi, \chi) \\ F(X, Y, \overleftarrow{Z}) &= \min(\Phi, \psi) - \min(\Phi, \psi, 1 - \chi) \\ F(X, \overleftarrow{Y}, Z) &= \min(\Phi, \chi) - \min(\Phi, 1 - \psi, \chi) \\ F(\overleftarrow{X}, Y, Z) &= \min(\psi, \chi) - \min(1 - \Phi, \psi, \chi) \end{aligned}$$

Dal confronto con la precedente [37] risulta subito che

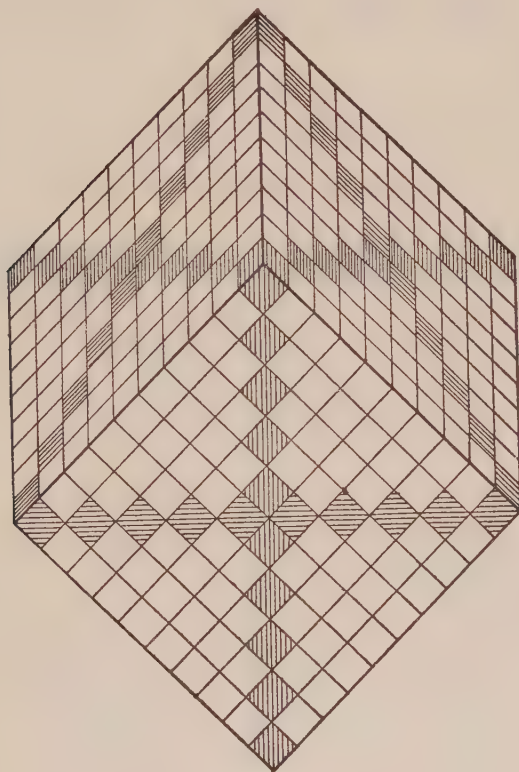


Fig. 9

(*) La freccia su un gruppo di 2-3 variabili significa cograduazione tra di esse; su una sola variabile significa contrograduazione della variabile stessa a ognuna delle altre, o al loro prodotto se il gruppo delle altre variabili è soprasedato con la freccia orientata in senso opposto.

$$F(\overrightarrow{X}, \overrightarrow{Y}, Z) = \max F(X, Y, Z) \quad (9) \quad [38]$$

mentre, poichè

$$F(X, Y, \overleftarrow{Z}) = 0 \text{ se } \min(\Phi + \chi - 1, \psi + \chi - 1) < 0$$

$$= \min(\Phi + \chi - 1, \psi + \chi - 1) \text{ se } \left. \begin{matrix} \Phi \\ \psi \end{matrix} \right\} > 1 - \chi$$

e quindi :

$$F(X, Y, \overleftarrow{Z}) = \max [0, \min(\Phi + \chi - 1, \psi + \chi - 1)]$$

e analogamente

$$F(X, \overleftarrow{Y}, Z) = \max [0, \min(\Phi + \psi - 1, \psi + \chi - 1)] \quad [39]$$

$$F(\overleftarrow{X}, Y, Z) = \max [0, \min(\Phi + \psi - 1, \psi + \chi - 1)]$$

Poichè le [38] e [39] corrispondono a distribuzioni lineari esse si otterranno facilmente dalle tabelle di co- e contrograduazione delle variabili triple, che si ottengono applicando due volte il procedimento di graduazione di Salvemini.

Nel nostro esempio, cograduiamo prima le variabili Y e Z :

$Z \backslash Y$	Y_1	Y_2	Y_3	Y_4	χ_i
Z_1	2				2
Z_2	2	2			4
Z_3			1	1	2
Z_4				2	2
ψ_j	4	2	1	3	10

(9) Questo metodo semplifica notevolmente il procedimento «pratico» indicato dal RIZZI (*Tabelle di congraduazione nel caso tridimensionale*, «Statistica», Anno XIX, 1, 1959).

Le caselle nelle quali le frequenze della variabile (Y, Z) risultano diverse da 0 costituiscono un insieme lineare e sono pertanto ordinabili a seconda del valore della somma degli indici delle modalità associate. Le coppie di modalità (Y_i, Z_j) associate a tali caselle costituiscono pertanto una variabile ordinata T_{i+j} che, cograduata con la variabile X dà la tabella di cograduazione della variabile tripla (X, Y, Z) :

$X \backslash T$	T_{1+1}	T_{1+2}	T_{2+2}	T_{2+3}	T_{3+3}	T_{3+4}	T_{4+4}	
X_1	1							1
X_2	1	2	2	—	1			6
X_3						1	2	6
	2	2	2	—	1	1	2	

cui corrisponde la funzione di ripartizione [38] mentre contrograduando alla stessa variabile X dà la tabella $(\overset{\leftarrow}{X}, Y, Z)$ di contrograduazione della variabile (X, Y, Z) :

$X \backslash T$	T_{1+1}	T_{1+2}	T_{2+2}	T_{2+3}	T_{3+3}	T_{3+4}	T_{4+4}	
X_1							1	1
X_2		1	2	—	1	1	1	6
X_3	2	1						3
	2	2	2	—	1	1	2	

la cui funzione di ripartizione è la terza delle [39].

Se modifichiamo ancora la metrica dello spazio rappresentativo della nostra variabile tripla in modo da rendere uguali le

ampiezze degli intervallini nei quali cade l'unità di massa su ciascun asse (che immagineremo concentrata nei punti di coordinate $i = 1, 2, \dots, n$, talchè le funzioni di ripartizione risultano tutte continue negli estremi inferiori di ciascun intervallo $i, i + 1$ e discontinue a destra), le variabili triple lineari della classe di Fréchet sono distribuite uniformemente lungo le 4 diagonali del cubo $O A B C D E F G$.

La diagonale $O E$ ha la funzione di ripartizione massima F_e ma le altre tre hanno funzioni di ripartizione distinte, per le quali si verifica dalle [37] e [39].

$$F_e(X, Y, Z) \geq \min F(X, Y, Z)$$

Con tale metrica sono costruite le due tabelle seguenti che danno i valori delle funzioni di ripartizione

$$F(\overrightarrow{X}, Y, Z) \text{ e } F(\overleftarrow{X}, \overrightarrow{Y}, Z) = F(\overrightarrow{Y}, Z) - F(\overleftarrow{X}, Y, Z)$$

$$F(\overrightarrow{X}, Y, Z) = \text{Max } F(X, Y, Z)$$

	Z_1		Z_2				Z_3		Z_4		
Y_4	I	2	3	4	5	6	7	8	9	10	Z_4
	I	2	3	4	5	6	7	8	9	9	
Y_3	I	2	3	4	5	6	7	8	8	8	Z_3
	I	2	3	4	5	6	7	7	7	7	
Y_2	I	2	3	4	5	6	6	6	6	6	Z_2
	I	2	3	4	5	5	5	5	5	5	
Y_1	I	2	3	4	4	4	4	4	4	4	Z_1
	I	2	3	3	3	3	3	3	3	3	
Y_1	I	2	2	2	2	2	2	2	2	2	Z_1
	I	I	I	I	I	I	I	I	I	I	
	X_1		X_2				X_3				

$F(\overleftarrow{X}, \overrightarrow{Y}, Z)$											
	X_3			X_2						X_1	
Y_4	1	2	3	4	5	6	7	8	9	10	X_1
	1	2	3	4	5	6	7	8	9	9	
	1	2	3	4	5	6	7	8	8	8	
Y_3	1	2	3	4	5	6	7	7	7	7	X_2
Y_2	1	2	3	4	5	6	6	6	6	6	
	1	2	3	4	5	5	5	5	5	5	
Y_1	1	2	3	4	4	4	4	4	4	4	X_3
	1	2	3	3	3	3	3	3	3	3	
	1	2	2	2	2	2	2	2	2	2	
	1	1	1	1	1	1	1	1	1	1	
	Z_1			Z_2		Z_3		Z_4			

In ciascuna delle due tabelle il valore di F da leggere è il massimo tra quelli scritti nella sezione delle colonne o delle righe corrispondenti alle due variabili riportate sugli assi, relativa alla zona interessata della terza variabile o alla prima che la precede nell'ordine considerato.

Così, per es., sia :

X, Y, Z	$F(X, Y, Z)$	$F(\overleftarrow{X}, \overrightarrow{Y}, Y)$	$F(\overrightarrow{Y}, Z)$	$F(\overleftarrow{X}, Y, Z)$
131	1	2	2	0
224	5	6	6	0
232	6	6	6	0
233	7	7	7	0
314	3	3	4	1
332	6	3	6	3

12. Ma esiste una distribuzione tripla per la quale sia in ogni punto

$$F(X, Y, Z) = \max(0, \Phi + \psi + \chi - 2)?$$

Noi abbiamo dimostrato che in ogni punto non può essere

$$F(X, Y, Z) < \max[0, \Phi + \psi + \chi - 2]$$

ma non che in effetti possa essere, in un punto determinato

$$F(X, Y, Z) = \max[0, \Phi + \psi + \chi - 2]$$

Dimostriamo ora che, qualunque sia $M(x_0, y_0, z_0)$, esiste sempre una distribuzione tripla della classe di Fréchet che soddisfa alla precedente.

Se ci riferiamo alla nostra rappresentazione modificata, nella quale è:

$$\Phi = \frac{x}{n}, \quad \psi = \frac{y}{n}, \quad \chi = \frac{z}{n},$$

la dimostrazione non perde nulla della sua generalità (10).

Consideriamo prima il caso

$$x + y + z - 2n \leq 0$$

e cioè

$$\Phi + \psi + \chi - 2 \leq 0.$$

Il punto M appartiene al solido $O A B C D E F G$ delimitato da quella parte delle facce del cubo rappresentativo che è dalla stessa banda dell'origine O rispetto al piano $x + y + z - 2n = 0$ e la sezione del cubo con detto piano.

Se indichiamo con M_x, M_y e M_z le proiezioni del punto M sui piani coordinati, almeno una di tali proiezioni appartiene ad uno dei triangoli rettangoli in O : $A O C, C O G, G O A$. Sia M_y appartenente a $G O C$. In tal caso è $n - z \geq x, n - x \geq z$ e quindi esistono almeno

$$\binom{n-z}{x} x! \binom{n-x}{z} z! (n-x-z)! n!$$

(10) L'osservazione si trova anche in FÉRON, già citato alla nota (1).

distribuzioni triple che non contengono unità nel parallelepipedo $O M_{xz} M_x M_{yz} M_y M_z M M_{xy}$ e per le quali è quindi $F(x, y, z) = 0$.

Tali distribuzioni si ottengono distribuendo prima le unità nel quadrato $O C D G$: x nel rettangolo $M_{xy} M_y R G$, scegliendo in esso x righe in una delle $\binom{n-z}{x}$ maniere possibili e disponendo queste x unità in una delle $x!$ maniere possibili. Procedendo analogamente nel rettangolo $M_{yz} C S M_y$ si ottengono $\binom{n-z}{x} x!$ $\binom{n-x}{z} z!$ distribuzioni di $x + z \leq n$ unità, e queste si completano disponendo nelle caselle intersezioni delle $n - x - z$ righe rimaste vuote con le $n - x - z$ colonne vuote le $n - x - z$ restanti. Ciò si può fare in $(n - x - z)$ maniere diverse. Ognuna delle distribuzioni così ottenute dà luogo a $n!$ distribuzioni diverse nel cubo, mediante traslazione parallela all'asse y .

Semplificando le precedenti espressioni si ottiene che il n° di tali distribuzioni è

$$\frac{(n-z)! (n-x)! n!}{(n-x-z)!} \quad \text{se} \quad y = n$$

Ma se $y < n$ alle disposizioni considerate sono anche da aggiungere le disposizioni che contengono fino a $\min(x, n-y)$ unità nel parallelepipedo $A K M_x M_{xz} M_x L I M$.

Le caselle lungo la GC costituiscono l'unica disposizione nella quale tutte le n unità siano al di fuori del triangolo GOC e quindi al di fuori del triedo $OACG$. Almeno per le $n!$ distribuzioni che si ricavano da queste è pertanto, in tutte le regioni $OACG$

$$F(x, y, z) = 0$$

Se il punto M appartiene al tetraedro $F B D E$ le sue proiezioni sui piani coordinati cadono nei triangoli $G C D$, $A C B$, $G A F$, ipotenuse escluse.

Operando come nel caso precedente si determinano sul piano $y = 0$

$$n - (n - z) - (n - x)$$

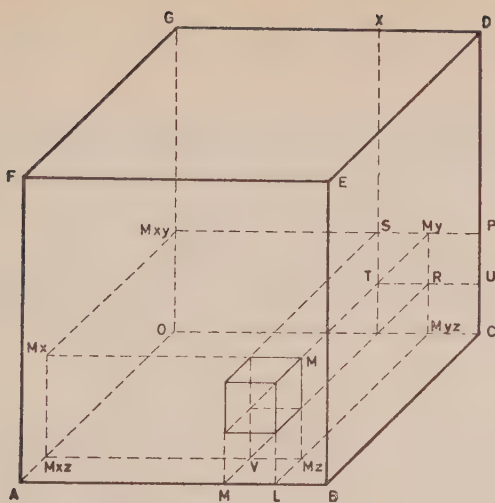


Fig. 12

righe e altrettante colonne appartenenti agli intervalli $0z$ e $0x$, nelle quali sono da situare altrettante unità. Nella figura tali intersezioni costituiscono il quadrato $STRMy$, ma nulla cambierebbe se esse fossero più o meno distanziate le une dalle altre. Le $n - z$ unità collocate entro $MxySKG$ e le $n - x$ collocate entro $MyzCUR$ sono esterne al parallelepipedo di diagonale OM : perchè anche le restanti $x + z - n$ possano collocarsi esternamente a detto parallelepipedo dovrebbero esse trovar posto nel parallelepipedo che si proietta sul piano $z = 0$ sul rettangolo $HLMyV$, per il che occorrerebbe fosse

$$n - y \geq x + z - n$$

e quindi

$$x + z + y \leq 2n$$

Poichè nel nostro caso è:

$$x_0 + y_0 + z_0 > 2n$$

solo

$$(n - x_0) + (n - z_0) + (n - y_0) = 3n - (x_0 + y_0 + z_0)$$

unità potranno trovare posto all'esterno del parallelepipedo di diagonale OM e le rimanenti :

$$n - [3n - (x_0 + y_0 + z_0)] = x_0 + y_0 + z_0 - 2n$$

dovranno essere poste nel parallelepipedo OM . Pertanto risulta

$$\min F(M) = \Phi + \psi + \chi - 2$$

Perchè la stessa distribuzione abbia funzione di ripartizione nulla in tutti i punti del cubo per i quali è

$$x + y + z - 2n \leq 0$$

occorre che le $k = x_0 + y_0 + z_0 - 2n$ unità interne al parallelepipedo di diagonale OM siano poste nei cubetti intorno al vertice M , che risultano sezionati dal piano $x + y + z = 2n$ o interamente da banda opposta all'origine O rispetto a tale piano. Queste sono 1 solo per $k = 1$, 4 per $k = 2$, 10 per $k = 3$ e in generale

$$\sum_1^k i \frac{i(i+1)}{2} = \frac{k(k+1)(k+2)}{6}$$

Pertanto la funzione di ripartizione di una stessa distribuzione tripla non può assumere il valore minimo: $\max(0, \Phi + \psi + \chi - 2)$ in tutti i punti del suo spazio rappresentativo, salvo che nel triedro $OACG$ nel quale una distribuzione può non avere nessun punto. Infatti una distribuzione la cui funzione di ripartizione dovesse essere nulla per tutti i punti per i quali è $x + y + z \leq 2n$ e uguale a 1 per tutti i punti per i quali $x + y + z - 2n = 1$ dovrebbe avere una unità in ciascuna delle $\frac{n(n+1)}{2}$ caselle aventi un vertice sul piano $x + y + z - 2n = 1$, e questo è impossibile. Se essa fosse quindi costantemente $= 1$ sul detto piano, dovrebbe necessariamente contenere unità al difuori del triedro $FBD E$ e avere funzione di ripartizione non nulla per $x + y + z - 2n < 0$.

Non esiste pertanto, nel caso di tre variabili, distribuzione estremante inferiore della classe di Fréchet, la cui funzione di ripartizione assume in tutti i punti il valore minimo possibile $\max(0, \Phi + \psi + \chi - 2)$.

Se una distribuzione ammette tale valore in un punto ne ammette certo uno più elevato in un altro.

13. Sorge la opportunità di definire in altra maniera la distribuzione estremante inferiore della classe di Fréchet e noi diremo che la distribuzione $v = f(x, y, z)$ è tale se l'integrale della sua funzione di ripartizione è un minimo.

Per le distribuzioni doppie tale definizione individua la linea di contrograduazione che ha come funzione di ripartizione la funzione di ripartizione minima

$$\min F = \max \{0, \Phi + \Phi - 1\}$$

mentre per le variabili triple o di ordine superiore individua un insieme più o meno ampio di distribuzioni.

L'integrale definito

$$I = \iiint F(X, Y, Z) \, dz \, dy \, dx$$

è facilmente calcolabile per le distribuzioni discontinue, e la suddivisione del campo in elementi di volume accoglienti masse unitarie semplifica all'estremo il problema.

Se una unità è localizzata nel punto x_0, y_0, z_0 essa porta contributo alla funzione di ripartizione in tutti i punti per i quali $x \geq x_0, y \geq y_0, z \geq z_0$ e quindi essa compare

$(n - x_0 + 1)(n - y_0 + 1)(n - z_0 + 1)$ volte nell'integrale I .

Pertanto

$$I = \sum_i^n (n - x_i + 1)(n - y_i + 1)(n - z_i + 1)$$

dove la somma è estesa agli n punti x_i, y_i, z_i nei quali sono localizzate le unità della distribuzione tripla.

I valori x_i costituiscono una permutazione dei numeri da 1 a n e similmente gli y_i e gli z_i . Se poniamo

$$x'_i = n - x_i + 1$$

$$y'_i = n - y_i + 1$$

$$z'_i = n - z_i + 1$$

si ottiene

$$I = \sum_i x'_i y'_i z'_i$$

Già sappiamo che I raggiunge il suo massimo quando le x_i le y_i e le z_i sono cograduate, e cioè che

$$\max I = \sum_1^n i^3 = \frac{n^2 (n + 1)^2}{4}$$

Perchè I raggiunga il suo massimo la condizione della cograduazione delle variabili deve essere verificata tra ciascuna variabile e il prodotto delle altre due. Una condizione analoga è necessaria che sia verificata per la distribuzione estremante inferiore e cioè che ciascuna variabile sia contrograduata al prodotto delle altre due.

A tale condizione non soddisfano le tre distribuzioni lineari di una classe tripla di Fréchet diverse dalla massimante, perchè in queste due delle variabili risultano cograduate e l'altra contrograduata con entrambe: i prodotti dei termini corrispondenti di due serie contrograduate non risultano contrograduati coi corrispondenti della terza serie, come nel seguente esempio ($n = 10$)

x'	y'	z'	$x' y' z'$
1	10	1	10
2	9	2	36
3	8	3	72
4	7	4	112
5	6	5	150
6	5	6	180
7	4	7	196
8	3	8	192
9	2	9	162
10	1	10	100
			1210

I numeri di ciascuna riga sono le coordinate dei punti nei quali sono collocate le unità delle distribuzioni rettilinee. Le altre due si ottengono permutando le variabili $x' y' z'$.

Le tre distribuzioni rettilinee ora considerate corrispondono al valore $I = 1210$, mentre la rettilinea massimante corrisponde al valore $\max I = 3025$.

Se cograduiamo la variabile z' col prodotto delle altre si ottiene :

x'	y'	z'
1	10	10
2	9	7
3	8	6
4	7	3
5	6	2
6	5	1
7	4	4
8	3	5
9	2	8
10	1	9

nella quale la y' non risulta cograduata col prodotto $x'z'$. Se eseguiamo tale cograduazione si ha :

x'	y'	z'
1	8	10
2	6	7
3	5	6
4	7	3
5	9	2
6	10	1
7	4	4
8	3	5
9	2	8
10	1	9

nella quale la z' non è cograduata con i prodotti $y'z'$: eseguiamo tali cograduazioni permutando le coppie $y'z'$:

x'	y'	z'
1	8	10
2	6	7
3	5	6
4	7	3
5	9	2
6	4	4
7	2	8
8	3	5
9	10	1
10	1	9

Tale procedimento ci porta alla costruzione delle variabili triple seguenti :

	<i>A</i>				<i>B</i>				<i>C</i>		<i>D</i>			
<i>x'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>
1	8	10	8	10	9	10	9	10	9	10	9	10	9	10
2	6	8	6	8	6	8	6	8	6	8	6	8	6	8
3	5	6	5	7	5	7	5	7	5	7	5	7	5	7
4	7	3	7	3	7	3	7	3	7	3	8	3	8	3
5	9	2	9	2	8	2	4	2	4	5	4	5	4	5
6	4	4	4	4	4	4	8	2	8	2	7	2	7	2
7	2	7	3	5	3	5	3	5	3	4	3	4	2	6
8	3	5	2	6	2	6	2	6	2	6	2	6	3	4
9	10	1	10	1	10	1	10	1	10	1	10	1	10	1
10	1	9	1	9	1	9	1	9	1	9	1	9	1	9
I	934		932		932		932		931		931		~931	

	<i>E</i>		<i>F</i>		<i>G</i>		<i>H</i>		<i>M</i>		<i>N</i>	
<i>x</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>	<i>y'</i>	<i>z'</i>
1	9	10	9	10	9	10	9	10	9	10	9	10
2	6	8	6	8	6	8	6	7	7	7	7	7
3	5	6	5	6	5	6	5	6	5	6	5	6
4	8	3	7	3	8	3	8	3	8	3	8	3
5	4	5	4	5	4	5	4	5	4	5	4	5
6	7	2	8	2	2	7	2	8	2	8	2	8
7	2	7	2	7	7	2	7	2	6	2	3	4
8	3	4	3	4	3	4	3	4	3	4	6	2
9	10	1	10	1	10	1	10	1	10	1	10	1
10	1	9	1	9	1	9	1	9	1	9	1	9
	930		930		930		930		930		930	

Ogni ulteriore permutazione dei valori delle 3 variabili che non aumenti I muta una delle distribuzioni *E F G H M N* in un'altra dello stesso gruppo. Ciò mostra che il risultato è raggiunto perchè nessun'altra distribuzione tripla può essere ottenuta nella quale una delle variabili non sia contrograduata col prodotto delle altre due, e quindi I non può essere ulteriormente ridotto. Le variabili triple estremanti si ottengono sostituendo le *x y z* corrispondenti alle *x', y' z'*. Esse sono pertanto :

	<i>E</i>		<i>F</i>		<i>G</i>		<i>H</i>		<i>M</i>		<i>N</i>	
<i>x</i>	<i>y</i>	<i>z</i>	<i>y</i>	<i>z</i>	<i>y</i>	<i>z</i>	<i>y</i>	<i>z</i>	<i>y</i>	<i>z</i>	<i>y</i>	<i>z</i>
1	10	2	10	2	10	2	10	2	10	2	10	2
2	1	10	1	10	1	10	1	10	1	10	1	10
10	2	1	2	1	2	1	2	1	2	1	2	1
3	8	7	8	7	8	7	8	7	8	7	5	9
4	9	4	9	4	4	9	4	9	5	9	8	7
5	4	9	3	9	9	4	9	3	9	3	9	3
6	7	6	7	6	7	6	7	6	7	6	7	6
7	3	8	4	8	3	8	3	8	3	8	3	8
8	6	5	6	5	6	5	6	5	6	5	6	5
9	5	3	5	3	5	3	5	4	4	4	4	4

Da ciascuna di tali distribuzioni se ne ottengono due permutando tra loro i valori delle tre variabili contenute nelle prime tre righe e 6 permutando i valori delle 7 righe successive.

Si hanno in tutto quindi $6 \times 2 \times 6 = 72$ distribuzioni estremanti inferiori della classe di Fréchet, nel senso che abbiamo definito sopra.

Nella fig. 13 sono rappresentate le posizioni sui piani coordinati della distribuzione *E*

Se al posto delle variabili semplici $\Phi(x)$ e $\psi(y)$ è data una variabile doppia $G(xy)$ e $G(x\infty) = \Phi(x)$, $G(\infty, y) = \psi(y)$, e la $\chi(z)$ è una sottoclasse di quella definita dalle Φ, ψ, χ , le sue distribuzioni estremanti si ottengono cograduando e contrograduando la variabile z al prodotto xy . Nella tabella sono date le due distribuzioni estremanti per una $G(xy)$ data, nella ipotesi che sia $\Phi = y$, $\psi = y$ e $\chi = z$:

<i>x</i>	<i>y</i>	<i>x' y'</i>	<i>z'_m = z_M</i>	<i>z'_M = z_m</i>	<i>x' y' z'_m</i>	<i>x' y' z'_M</i>
10	6	5	10	1	50	50
1	10	10	9	2	90	20
5	9	12	8	3	96	36
9	3	16	7	4	112	64
6	7	20	6	5	120	100
4	8	21	5	6	105	126
7	4	28	4	7	112	196
8	1	30	3	8	90	240
2	5	54	2	9	108	486
3	2	72	1	10	72	720
					955	2038

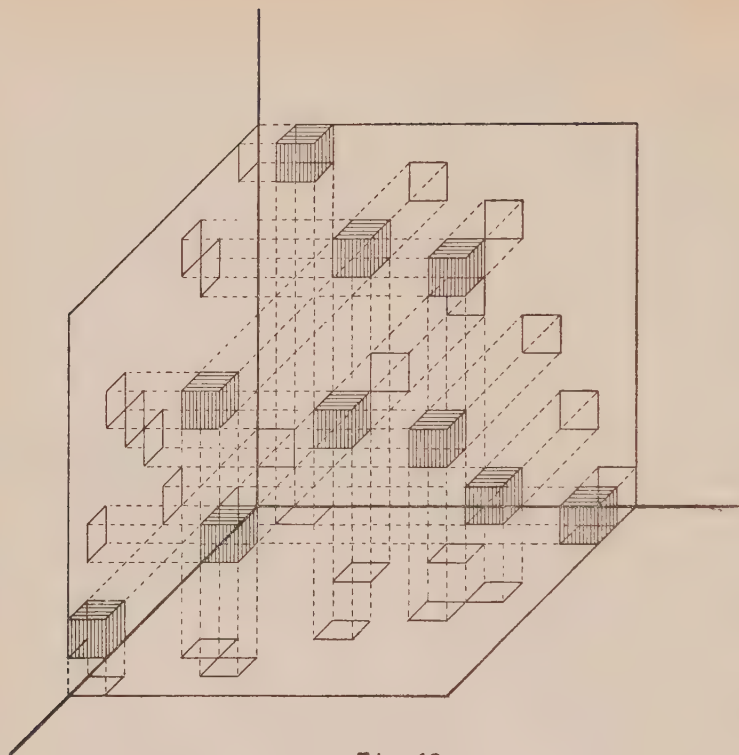


Fig. 13

Entrambe tali distribuzioni sono estremanti nel senso della nostra definizione, perchè corrispondono ai valori minimi e massimi dell'integrale della funzione di ripartizione compatibile coi vincoli della classe.

Gli estremi della $F(X, Y, Z)$ nei singoli punti sono :

$$\max F(X, Y, Z) = \min(\chi, G)$$

$$\min F(X, Y, Z) = G(x, y) - \max F = G(x, y) (1 - \chi, G)$$

e quindi :

$$\begin{aligned} \min F(X, Y, Z) &= 0 && \text{se } 1 - \chi > G \supset G + \chi - 1 < 0 \\ &= G + \chi - 1 && \text{se } 1 - \chi < G \supset G + \chi - 1 > 0 \end{aligned}$$

e, in definitiva

$$\max(0, G + \chi - 1) \leq F(X, Y, Z) \leq \min(G, \chi)$$

Perchè la funzione $F(X, Y)$ sia $= 1/n$ in ogni punto nel quale $G(x, y) = 1/n$ bisogna che la distribuzione tripla abbia una unità nel punto $(1, 1, 1)$ e in generale, perchè sia

$$F(X, Y, i) = i/n \text{ in ogni punto nel quale } G(x, y) = i/n$$

occorre che la distribuzione abbia 1 unità nei punti $(1, 1, 1)$, $(2, 2, 2), \dots, (i, i, i)$ e che quindi $G(x, y) = \min(\Phi(x), \psi(y))$ e la sottoclasse (G, χ) ha la stessa distribuzione estremante superiore della classe più vasta (Φ, ψ, χ) .

Perchè una distribuzione tripla della classe (G, χ) abbia come funzione di ripartizione la $F(X, Y, Z) = \max(0, G + \chi - 1)$,

poichè $\frac{z}{n} = \chi$, deve essere $F(X, Y, 1) = 0$, per $G \leq \frac{(n-1)/n}{1}$

e quindi la distribuzione deve avere una unità nel vertice (n, n, n)

e dev'essere $F(X, Y, 2) = \frac{0}{2/n}$ per $G \leq \frac{(n-2)/n}{(n-1)/n} = \frac{1}{1}$ e quindi la di-

stribuzione deve avere una seconda unità nel vertice $(n-1, n-1, n-1)$, e così via.

Pertanto tra le sottoclassi $[G, \chi]$ della classe (Φ, ψ, χ) solo la $G = \Phi = \psi$ ammette una funzione di ripartizione minima e una funzione di ripartizione massima.

14. In generale, possiamo quindi enunciare il teorema: Una classe di Fréchet delle distribuzioni triple (X, Y, Z) definite dalla funzione di ripartizione delle variabili X, Y, Z (Φ, ψ, χ) ammette un estremo superiore, costituito dalla distribuzione la cui funzione di ripartizione assume in ogni punto del campo di definizione il valore massimo compatibile coi vincoli, e un insieme di estremanti inferiori per tutte le distribuzioni, per le quali l'integrale delle funzioni di ripartizione è minimo.

Le variabili triple definite dalle funzioni di ripartizione di una distribuzione doppia della classe precedente e dalla restante di-

sistribuzione semplice costituiscono sottoclassi delle (Φ, ψ, χ) , che non ammettono in generale nè un estremo superiore nè un estremo inferiore, ma solo un insieme estremante superiore e un insieme estremante inferiore di distribuzione, l'integrale delle cui funzioni di ripartizione è rispettivamente massimo e minimo.

Solo se la distribuzione doppia è l'estremo superiore di una delle tre classi di Fréchet di distribuzioni doppie contenute nella classe tripla $(\Phi = \psi, \psi = \chi, \chi = \Phi)$ allora la sottoclasse tripla corrispondente ha per estremo superiore l'estremo superiore della classe (Φ, ψ, χ) e come estremo inferiore una delle altre tre distribuzioni lineari di detta classe.

D.P. BANERJIE

On the forms of some invariants of probability distribution

Mahalanobis (1936) first introduced the concept of distance in statistics. Jeffreys (1946, 1948) defined some invariants of probability distribution to measure the distance. Here I shall find the exact form of the invariants defined by Jeffreys to represent distance between two contiguous parametric points.

Let p and p' be the density functions of two probability distributions of a variate x . Then the invariants defined by Jeffreys are

$$I_m = \int \left| p'^{\frac{1}{m}} - p^{\frac{1}{m}} \right|^m dx$$

$$J = \int (p' - p) \log \frac{p'}{p} dx$$

$$\text{Let } p = \varphi(x, \alpha_1, \alpha_2 \dots \alpha_p),$$

$$p' = \varphi(x, \alpha_1 + d\alpha_1, \alpha_2 + d\alpha_2 \dots \alpha_p + d\alpha_p)$$

$$= \varphi(x, \alpha_1 \alpha_2 \dots \alpha_p) + \sum \frac{\partial \varphi}{\partial \alpha_\mu} d\alpha_\mu + \dots$$

by Taylor's theorem.

$$J = \int \left[\sum \frac{\partial \varphi}{\partial \alpha_\mu} d\alpha_\mu \right] \log \left[1 + \sum \frac{\partial \log \varphi}{\partial \alpha_\mu} d\alpha_\mu \right]$$

neglecting small quantities of higher order.

Now

$$\int \frac{\partial \log \varphi}{\partial \alpha_\mu} \varphi dx = \frac{\partial}{\partial \alpha_\mu} \int \varphi dx = 0 \quad (1)$$

since

$$\int \varphi dx = 1$$

Differentiating (1) with respect to α_ν we have

$$\int \frac{\partial^2 \log \varphi}{\partial \alpha_\mu \partial \alpha_\nu} \varphi dx + \int \frac{\partial \log \varphi}{\partial \alpha_\mu} \frac{\partial \varphi}{\partial \alpha_\nu} dx = 0$$

Hence

$$\int \frac{\partial \log \varphi}{\partial \alpha_\mu} \frac{\partial \varphi}{\partial \alpha_\nu} dx = - \int \frac{\partial^2 \log \varphi}{\partial \alpha_\mu \partial \alpha_\nu} \varphi dx \dots \quad (2)$$

$$= E \left(- \frac{\partial^2 \log \varphi}{\partial \alpha_\mu \partial \alpha_\nu} \right) = E_{\mu\nu} = E_{\nu\mu}$$

Hence

$$\begin{aligned} J &= \sum_{\mu=1}^p \sum_{\nu=1}^p E_{\mu\nu} d\alpha_\mu d\alpha_\nu \\ &= \sum \sum g_{\mu\nu} d\alpha_\mu d\alpha_\nu \dots \end{aligned} \quad (3)$$

where

$$g_{\mu\nu} = E_{\mu\nu} = g_{\nu\mu}$$

Rao (1945) has shown that the distance between two contiguous points is given by $ds^2 = \sum \sum \partial_{\mu\nu} d\alpha_\mu d\alpha_\nu$ where ds = distance between two contiguous parametric points.

Hence

$$s = \int_{t_1}^{t_2} \sqrt{\sum \sum g_{\mu\nu} \frac{d\alpha_\mu}{dt} \frac{d\alpha_\nu}{dt}} dt$$

if α_μ, α_ν are functions of t .

Now

$$\begin{aligned} I_2 &= \int (p'^{\frac{1}{2}} - p^{\frac{1}{2}})^2 dx = 2 - 2 \int \sqrt{p'p} dx \\ &= 2 - 2 \int \varphi \left[1 + \sum \frac{\partial \log \varphi}{\partial \alpha_\mu} d\alpha_\mu + \frac{1}{2\varphi} \sum \sum \frac{\partial^2 \varphi}{\partial \alpha_\mu \partial \alpha_\nu} d\alpha_\mu d\alpha_\nu \right]^{\frac{1}{2}} dx \\ &= 2 - 2 \int \varphi \left[1 + \frac{1}{2} \sum \frac{\partial \log \varphi}{\partial \alpha_\mu} d\alpha_\mu + \frac{1}{4\varphi} \sum \sum \frac{\partial^2 \varphi}{\partial \alpha_\mu \partial \alpha_\nu} d\alpha_\mu d\alpha_\nu \right] \end{aligned}$$

$$\begin{aligned}
 & -\frac{1}{8} \cdot \frac{1}{\varphi^2} \left\{ \Sigma \frac{\partial \varphi}{\partial \alpha_\mu} d \alpha_\mu \right\}^2 d x \\
 = & -\frac{1}{2} \int \varphi \left[\frac{1}{\varphi} \Sigma \Sigma \frac{\partial^2 \varphi}{\partial \alpha_\mu \partial \alpha_\nu} d \alpha_\mu d \alpha_\nu - \frac{1}{2} \left\{ \frac{1}{\varphi} \Sigma \frac{\partial \varphi}{\partial \alpha_\mu} d \alpha_\mu \right\}^2 \right] d x
 \end{aligned}$$

Since

$$\int \frac{\partial \varphi}{\partial \alpha_\mu} d x = 0$$

Now

$$\frac{\partial^2 \log \varphi}{\partial \alpha_\mu \partial \alpha_\nu} = \frac{1}{\varphi} \frac{\partial^2 \varphi}{\partial \alpha_\mu \partial \alpha_\nu} - \frac{1}{\varphi^2} \frac{\partial \varphi}{\partial \alpha_\mu} \frac{d \varphi}{\partial \alpha_\nu} \dots \quad (4)$$

Again

$$\int \frac{\partial \log \varphi}{\partial \alpha_\mu} \frac{\partial \log \varphi}{\partial \alpha_\nu} \varphi d x = - \int \frac{\partial^2 \log \varphi}{\partial \alpha_\mu \partial \alpha_\nu} \varphi d x$$

by (2) Hence

$$I_2 = -\frac{1}{4} \Sigma \Sigma E_{\mu\nu} d \alpha_\mu d \alpha_\nu = \frac{J}{4} \dots \quad (5)$$

Illustrative Example.

Let us consider Koopman's (1936) general form of distribution admitting sufficient statistics

$$f(x, \alpha_1, \alpha_2, \dots, \alpha_p) = e x p \left[\sum_{k=1}^p u_k v_k(x) + A(x) + B \right]$$

where u'_k 's are functions of $\alpha_1, \alpha_2, \dots, \alpha_p$. B is also function of $\alpha_1 \alpha_2 \dots \alpha_p$. Hence we have $f(x, \alpha_1, \alpha_2, \dots, \alpha_p) = f_1(x, B, u_1, u_2, \dots, u_p)$ assuming $B, \alpha_1, \alpha_2, \dots, \alpha_p$ as functions of u_1, u_2, \dots, u_p . Hence

$$J = - \Sigma \Sigma \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} d u_k d u_{k_1} \dots \quad (6)$$

Huzurbazar (1955) proved that

$$I_2 = 2 - 2 e x p \left[\frac{B + B'}{2} - b \left(\frac{u_k + u'_k}{2} \right) \right]$$

Where B' 's are functions of u'_k 's and $B = b(u_k)$

Let $u'_k = u_k + \delta u_k$. Then

$$\begin{aligned} B' &= B + \Sigma \frac{\partial B}{\partial u_k} d u_k + \frac{1}{2} \Sigma \Sigma \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} d u_k d u_{k_1} + \\ &b\left(\frac{u_k + u'_k}{2}\right) = b\left(u_k + \frac{1}{2} \delta u_k\right) = B + \frac{1}{2} \Sigma \frac{\partial B}{\partial u_k} d u_k \\ &+ \frac{1}{8} \Sigma \Sigma \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} d u_k d u_{k_1} \end{aligned}$$

Hence

$$\begin{aligned} \frac{B + B'}{2} - b\left(\frac{u_k + u'_k}{2}\right) &= \frac{1}{8} \Sigma \Sigma \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} d u_k d u_{k_1} \\ I_2 &= 2 - 2 \exp \left[\frac{1}{8} \Sigma \Sigma \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} d u_k d u_{k_1} \right] \\ &= -\frac{1}{4} \Sigma \Sigma \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} d u_k d u_{k_1} = \frac{J}{4} \dots \end{aligned} \quad (7)$$

Let us consider Beta variate population given by

$$\varphi(x, p, q) = \frac{x^{p-1} (1-x)^{q-1}}{B(p, q)}, \quad 0 \leq x \leq 1$$

$$\begin{aligned} I_2 &= 2 - 2 \int_0^1 \frac{x^{\frac{p+p'}{2}-1} (1-x)^{\frac{q+q'}{2}-1}}{\sqrt{B(p, q) B(p', q')}} dx \\ &= 2 - 2 \frac{B\left(\frac{p+p'}{2}, \frac{q+q'}{2}\right)}{\sqrt{B(p, q) B(p', q')}} \end{aligned} \quad (8)$$

$$J = \int_0^1 dx \left\{ \frac{x^{p'-1} (1-x)^{q'-1}}{B(p', q')} - \frac{x^{p-1} (1-x)^{q-1}}{B(p, q)} \right\} \times$$

$$\begin{aligned}
& \left\{ -\log \frac{B(p', q')}{B(p, q)} + (p' - p) \log x + (q' - q) \log (1 - x) \right\} \\
&= \delta p \frac{\delta}{\delta p} \log \frac{B(p + \delta p, q + \delta q)}{B(p, q)} + \delta q \frac{d}{\delta q} \log \frac{B(p + \delta p, q + \delta q)}{B(p, q)} \\
&= \frac{\delta}{\delta p} \log \frac{B + \frac{\delta B}{\delta p} d p + \frac{\delta B}{\delta q} d q}{B} d p \\
&\quad + \frac{\delta}{\delta q} \log \frac{B + \frac{\delta B}{\delta p} d p + \frac{\delta B}{\delta q} d q}{B} d q \\
&= \Sigma \Sigma \frac{\delta^2 \log B}{\delta \mu \delta \nu} d \mu d \nu \quad \text{where} \quad \begin{matrix} \mu = p, q \\ \nu = p, q \end{matrix} \\
&= - \Sigma \Sigma E \left(\frac{\partial^2 \log B}{\partial \mu \partial \nu} \right) d \mu d \nu \dots \tag{10}
\end{aligned}$$

by putting

$$p' = p + \delta p, q' = q + \delta q$$

Similarly

$$\begin{aligned}
I_2 &= 2 - 2 \frac{B\left(p + \frac{\delta p}{2}, q + \frac{\delta q}{2}\right)}{\sqrt{B} \sqrt{B + \frac{\delta B}{\delta p} d p + \frac{\delta B}{\delta q} d q}} \\
&= \frac{1}{4} \Sigma \Sigma \frac{\partial^2 \log B}{\delta \mu \delta \nu} d \mu d \nu = \frac{J}{4} \dots \tag{11}
\end{aligned}$$

Let $\beta_1, \beta_2, \dots, \beta_p$ be orthogonal parameters.

Let $\alpha_\mu = f_\mu(\beta_1, \beta_2, \dots, \beta_p)$ and hence $\beta_\mu = F_\mu(\alpha_1, \alpha_2, \dots, \alpha_p)$

And

$$\varphi(x, \alpha_1, \alpha_2, \dots, \alpha_p) = \varphi_1(x, \beta_1, \beta_2, \beta_p)$$

Now (Jeffrey 1948) the parameters β_μ 's are said to be orthogonal of

$$E \left\{ -\frac{\partial^2 \log \varphi_1}{\partial \beta_\mu \partial \beta_\nu} \right\} = 0 \quad \text{when} \quad (\mu \neq \nu)$$

Hence

$$J = \Sigma E \left(-\frac{\partial^2 \log \varphi}{\partial \beta_\mu^2} \right) (d\beta_\mu)^2 = \Sigma g_{\mu\mu} (d\beta_\mu)^2$$

Conversely if $J = \Sigma g'_{\mu\mu} (d\beta_\mu)^2$ then $g'_{\mu\nu} = 0$ ($\mu \neq \nu$) [Weatherburn 1950, p. 44].

Hence $E \left[-\frac{\partial^2 \log \varphi}{\partial \beta_\mu \partial \beta_\nu} \right] = 0$. Hence are β_μ 's orthogonal coordinates or parameters.

Example. Consider Pearson's Type III distribution.

Let

$$\begin{aligned} \varphi(x, a, p) &= \frac{a^p e^{-ax} x^{p-1}}{\Gamma_p} \quad (0 \leq x \leq a) \\ &= e^{-ax+p \log x - \log x + p \log a - \log \Gamma(p)} \end{aligned}$$

Hazurbarar (1956) proved that $\beta_1 = \frac{p}{a}$, $\beta_2 = p$ are orthogonal coordinates.

Hence

$$\begin{aligned} \log \varphi(x, \beta_1, \beta_2) &= -\frac{\beta_2}{\beta_1} x + \beta_2 \log x - \log x \\ &\quad + \beta_2 [\log \beta_2 - \log \beta_1] - \log \Gamma(\beta_2) \end{aligned}$$

Hence

$$\begin{aligned} 4 I_2 = J &= \left[\frac{1}{\beta_2} - \frac{d}{d\beta_2} \frac{\Gamma'(\beta_2)}{\Gamma(\beta_2)} \right] (d\beta_2)^2 \\ &\quad + \frac{\beta_2}{\beta_1^2} \left[1 - \frac{2}{\beta_1} \right] (d\beta_1)^2 \end{aligned}$$

Equation (3) denotes fundamental tensor in Riemannian space V_p of p dimensions. (Weatherburn p. 44).

As an arbitrary Riemannian V_p does not admit an n -ply orthogonal system of hypersurfaces it is not possible to find orthogonal coordinates if $p > 3$ when the fundamental form is arbitrary.

As (Weatherburn p. 51) it is possible to immerse Riemannian space V_p in Euclidean space s_m of m dimensions where

$m > \frac{p(p+1)}{2}$ it is possible to express the fundamental form in Euclidean coordinates and hence in orthogonal Euclidean coordinates.

The parametric differential equation of geodesics (Weatherburn p 75) will be

$$\frac{\partial^2 \alpha_2}{ds^2} + \left\{ \begin{matrix} l \\ jk \end{matrix} \right\} \frac{\partial \alpha_i}{ds} \frac{d \alpha_k}{ds} = 0$$

where

$$\left\{ \begin{matrix} l \\ jk \end{matrix} \right\} = \frac{g^{kh}}{2} \left[\frac{\partial g_{jk}}{\partial \theta_i} + \frac{\partial g_{ih}}{\partial \theta_j} - \frac{\partial g_{ij}}{\partial \theta_k} \right]$$

where $g^{kh} g^{hl} = \delta_l^k = 1$ or 0 according as $k = l$ or hence the parametric differential equation of geodesics will be

$$\frac{\partial^2 \alpha_l}{ds^2} + \frac{E^{lk}}{2} \left[\frac{\partial E_{kh}}{\partial \alpha_i} + \frac{\partial E_{ih}}{\partial \alpha_k} - \frac{\partial E_{ik}}{\partial \alpha_h} \right] \frac{d \alpha_i}{ds} \frac{d \alpha_k}{ds} = 0$$

where

$$\begin{aligned} E^{kh} E_{hl} &= \delta_l^k = 1 \text{ if } k = l \\ &= 0 \text{ if } k \neq l. \end{aligned}$$

Numerical Examples.

Example 1

Putting

$$B = u_1^2 + u_2^2 + u_3^2 + u_4^2$$

in the result

$$J = - \sum \sum \frac{\partial^2 B}{\partial u_k \partial u_{k_1}} \frac{d u_k}{ds} \frac{d u_{k_1}}{ds}$$

we have

$$J = -2 [(du_1)^2 + (du_2)^2 + (du_3)^2 + (du_4)^2]$$

Example 2.

Putting

$$p = 2, q = 2$$

in the Beta variate population

$$I = .36 [(d p)^2 + (d q)^2] - .34 \delta p \delta q$$

Example 3.

Putting

$$p = 2, q = 2$$

in Pearson's Type III distribution

$$\beta_2 = 2, \beta_1 = 1$$

$$J = -.03 (d \beta_2)^2 - 2 (d \beta_1)^2.$$

References

- HUZURBAZAR V. S. (1955) *Exact form of some invariants for distributions admitting sufficient statistics*. « *Biometrika* », 42, 533.
- (1956) *Sufficient statistics and orthogonal parameters* « *Sankhya* », 12, 217.
- JEFFREYS, H. (1946) *An invariant form for the prior probability in estimation problems*, Proc. Roy Soc. A 186, 453.
- (1948) *Theory of Probability*, 2nd ed. Oxford Clarendon Press.
- KOOPMAN, B.O. (1936) *On distribution admitting a set of sufficient statistics*. « *Trans Amer Math. Soc.* », 39, 399.
- MAHALANOBIS P. C. (1938). *On the generalized distance in statistics*, « *Proc. Nat. Inst. Soc.* », India, 12.
- RAO, C. R. (1945) *Information and accuracy attainable in estimation of statistical parameters* « *Brulat Math. Soc.* », 37, 81.
- WEATHERBURN (1950). *Rieman Geometry and Tensor Calculus*, Isted Cambridge, pp. 50, 51, 44.
- S. V. UNIVERSITY. TIRUPATI ANDHRA PRADESH, INDIA.

CARLO BENEDETTI

Su alcune singolari questioni statistiche ed econometriche connesse con le medie

1. In alcune questioni statistiche si presentano delle disuguaglianze fra valori medi che a prima vista possono indurre a perplessità circa l'esattezza dei calcoli effettuati in quanto sembrano sospette o addirittura incompatibili con i dati di partenza.

Ad esempio, dal censimento effettuato dall'ISTAT nel 1954 sui dipendenti delle Amministrazioni Statali (1), si ha il seguente prospetto :

	Di ruolo		Avventizi		Complesso	
	stipendio medio (lire)	numero	stipendio medio (lire)	numero	stipendio medio (lire)	numero
Insegnanti di scuola elementare .	52.343	154.671	35.988	18.506	50.596	173.177
Insegnanti di scuola media . . .	64.904	34.081	37.831	57.836	47.869	91.917

Dove si nota che lo stipendio medio degli insegnanti di scuola media è superiore a quello dei maestri elementari sia fra gli insegnanti di ruolo che fra gli avventizi, mentre, viceversa, nel complesso i maestri elementari presentano uno stipendio medio più alto degli insegnanti di scuola media.

(1) *Rilevazione Statistica dei Dipendenti delle Amministrazioni Statali* in « Notiziario ISTAT » Serie C, suppl. al n. 11 ; 23 dicembre 1955 ed elaborazioni della rivista « *Rassegna di Statistiche del Lavoro* » n. 6, nov.-dic. 1955, pag. 527.

Che la cosa sia possibile e che, anzi, nel caso illustrato del prospetto, si tratti di risultato giusto alieno da errori di calcolo non vi sono dubbi. Accenneremo qui di seguito a semplici considerazioni a chiarimento di quanto sopra e di questioni analoghe.

2. Allo scopo di illustrare il caso su esposto ed altri casi analoghi si rende opportuno stabilire una proposizione generale che comprenda i suddetti casi. Inoltre, il teorema che dimostriamo consente di stabilire per due sistemi di pesi le condizioni *necessarie e sufficienti* affinchè una media ponderata sia non inferiore ad un'altra, mentre, se non sbaglio, finora si conoscono o, se si vuole, si sono scritte solo condizioni sufficienti.

Ciò è reso possibile in quanto si considera una media ponderata di due soli termini a, b con pesi $\lambda, 1 - \lambda$ ($0 \leq \lambda \leq 1$):

$$m = a \lambda + (1 - \lambda) b$$

forma alla quale può essere ricondotta ogni media ponderata con qualunque numero di termini; basta cioè, aggregare tutti i termini in due medie ponderate parziali a e b e quindi considerare la media ponderata di queste con i pesi λ ed $1 - \lambda$ che le competono.

L'avere ridotto il problema a due soli termini consente di ridurre notevolmente la casistica che scaturisce dal numero dei termini e che cresce a dismisura divenendo intrattabile al crescere di questo numero. Tale riduzione permette quindi di aderire alle esigenze del problema di scrivere anche le condizioni *necessarie* per i due soli pesi λ ed λ' rispettivamente delle medie:

$$m = a \lambda + (1 - \lambda) b$$

ed

$$m' = a' \lambda' + (1 - \lambda') b'$$

3. Sia $m = a \lambda + (1 - \lambda) b$; $m' = a' \lambda' + (1 - \lambda') b$ con $a, b, a', b', \lambda, \lambda'$ reali di cui $a > b$ e $0 \leq \lambda, \lambda' \leq 1$; ora, ponendo $\frac{a' - b'}{a - b} = c$; $\frac{b' - b}{a - b} = d$ si ha il teorema:

TEOREMA : Condizione necessaria e sufficiente affinchè si abbia :

$$m \geq m' \quad (1)$$

è che si verifichi

$$\max (0, c \lambda' + d) \leq \lambda \leq 1 \quad (2)$$

DIMOSTRAZIONE : La (2) è necessaria : infatti se :

$$m = a \lambda + (1 - \lambda) b \geq m' = a' \lambda' + (1 - \lambda') b'$$

è :

$$\lambda \geq \frac{a' - b'}{a - b} \lambda' + \frac{b' - b}{a - b} = c \lambda' + d$$

ora, considerando che λ non può essere esterno all'intervallo chiuso $(0, 1)$ e l'ambiguità di segno di $c \lambda' + d$, è :

$$1 \geq \lambda \geq \max (0, c \lambda' + d)$$

La (2) è sufficiente : ricordando che $y = a x + (1 - x) b = (a - b) x + b$ cresce con x in $(0, 1)$ se $a - b > 0$ e ricordando le espressioni di c e di d si ha :

$$\begin{aligned} m &= (a - b) \lambda + b \geq (a - b) \max (0, c \lambda' + d) + \\ &+ b = \max (b, m') \geq m'. \end{aligned}$$

C.D.D.

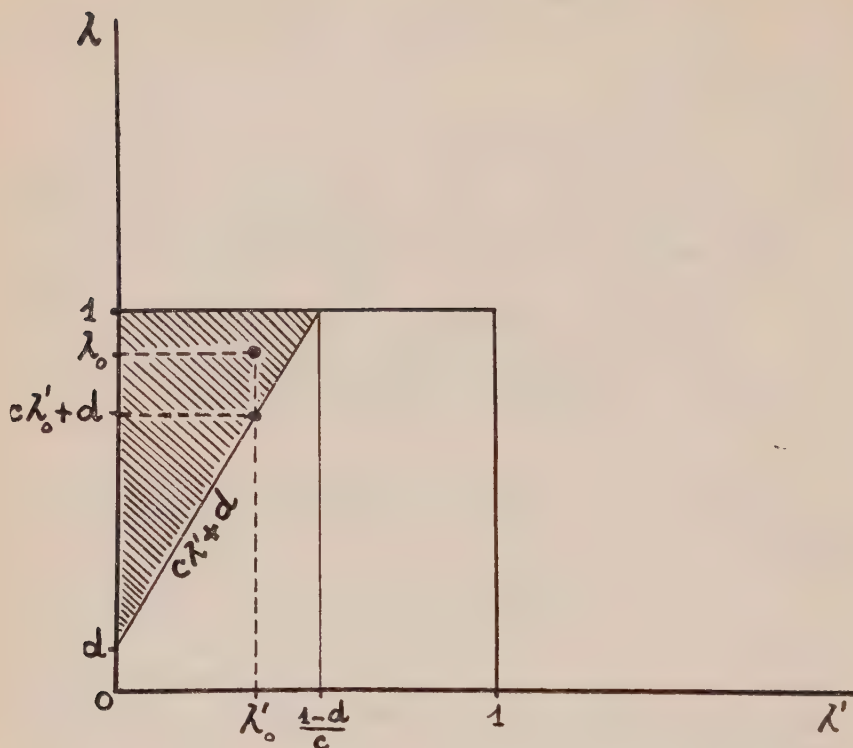
4. Il caso dello stipendio medio dei professori di scuola media e dei maestri elementari accennato nel n. 1 rientra nel teorema come si vede immediatamente :

$$a = 52343 ; b = 35988 \quad ; a' = 64904 \quad ; b' = 37831$$

$$\lambda = 0,893138 ; \lambda' = 0,37078 \quad ; m = 50596 \quad ; m' = 47869$$

$$0 < c \lambda' + d = 0,7264445 < \lambda < 1 ; c = 1,65533 ; d = 0,1127.$$

Semplicissima ne è l'illustrazione geometrica :



La superficie tratteggiata compreso il contorno rappresenta il dominio dei punti (λ, λ') che determinano, dati i suddetti valori di a, b, a', b' , la disuguaglianza :

$$m \geq m'$$

Quest'area potrebbe interpretarsi come la probabilità che, dati i suddetti valori di a, b, a', b' , si abbia $m \geq m'$ qualora le coppie (λ, λ') fossero ugualmente possibili (avessero cioè densità uniforme) nel quadratino di lato uno.

5. Nel caso delle medie ponderate m ed m' di due successioni di termini non tutti uguali, rispettivamente a_1, a_2, \dots, a_n ed a'_1, a'_2, \dots, a'_n , ($n > 2$) è chiaro che si può avere ancora $m \geq m'$

pur essendo $a_i \leq a'_i$, per tutti gli i da 1 ad n , cioè lo stesso caso, sebbene generalizzato, degli stipendi medi degli insegnanti elementari e medi.

Intanto si comincia col notare che occorre, per avere $m \geq m'$, almeno una $a_i \geq a'_i$ ($i \neq j$) ⁽²⁾; ciò si verifichi, ad es., per $\max_i a_i$, in tal caso m ed m' possono scriversi nel modo seguente dove $a = \max_i a_i \geq a'_i$, ($i \neq j$):

$$m = a \lambda + (1 - \lambda) b$$

$$m' = a'_j \lambda' + (1 - \lambda') b'$$

e dove λ è il peso di a ; b è la media ponderata dei rimanenti $n - 1$ termini di m ; analogo significato hanno λ' e b' .

Ora, essendo per ipotesi $a > b$ (si ricordi che $a = \max_i a_i$ e le a_i non sono tutte uguali) è chiaro che si avrà $m \geq m'$ se e solo se λ e λ' soddisfano la (2) del Teorema.

6. Nel campo commerciale casi formalmente analoghi a quello degli stipendi degli insegnanti medi ed elementari sono frequenti. Basti pensare a due cinema A ed A' di cui A pratici i prezzi di biglietto a e b rispettivamente di galleria e di platea ed A' pratici i prezzi a' e b' . Nel caso di A ed A' al completo, è ovvio che se $a < a'$ e $b < b'$, essendo per ipotesi $a > b$ ed $a' > b'$, qualora sia $a > b'$ ⁽³⁾ si avrà:

$$m = a \lambda + (1 - \lambda) b > m' = a' \lambda' + (1 - \lambda) b'$$

se la percentuale λ dei posti di galleria di A e l'analogha percentuale λ' di A' soddisfano alla condizione (2) del Teorema; si avrà, cioè in tal caso, che il prezzo medio m del biglietto di A risulta più alto del prezzo medio m' di A' malgrado sia la platea che la galleria costino meno nel cinema A che nel cinema A' .

(2) Altrimenti, come è ovvio, siccome sarebbe: $\max_i a_i < \min_i a'_i$, ed essendo $\min_i a_i < m < \max_i a_i$ è chiaro che sarebbe $m < m'$.

(3) Altrimenti, ovviamente, si avrebbe $\max(a, b) < \min(a', b')$ e quindi sarebbe comunque: $m < m'$.

Lo stesso può dirsi a proposito di due ditte A ed A' che vendano ambedue lo stesso prodotto in due piazze; e di numerosi altri analoghi esempi.

Sempre rimanendo in questo tipo di applicazioni di carattere economico, si può dire che affinchè queste disuguaglianze possano fornire criteri di condotta in una competizione fra operatori economici, fra imprese, ecc., occorre, evidentemente, tenere conto della legge della domanda dei prodotti, anche se per alcuni prodotti a « consumo rigido » si può prescindere da tale legge in quanto la manovra delle quantità dei prodotti immessi nel mercato non altera o quasi il prezzo quando sia contenuta entro certi limiti.

Accenniamo al caso in cui nelle medie: $m' = \lambda' a' + (1 - \lambda') b'$; $m = \lambda a + (1 - \lambda) b$ le lettere: a', b', a, b indichino i prezzi di uno stesso prodotto offerto da A' ed A nelle stesse due piazze o settori; più precisamente indichiamo con Q' il volume di produzione di A' e con Q il volume di produzione di A da destinarsi alla vendita nelle suddette due piazze o settori; inoltre per la legge della domanda sia nella prima piazza o settore rispettivamente per A' ed A :

$$a' = f_1(\lambda' Q'); \quad a = f_1(\lambda Q)$$

e nella seconda piazza o settore rispettivamente per A' ed A :

$$b' = f_2[(1 - \lambda') Q']; \quad b = f_2[(1 - \lambda) Q]$$

Si richiederà solo, per la stessa definizione di legge di domanda, che $f_1(x)$, $f_2(x)$ per tutti gli x che interessa considerare, siano finite e decrescenti.

Ora se, ad es., A vuole fare concorrenza ad A' , secondo il nostro criterio, praticherà prezzi più bassi in ciascuna delle due piazze cercando di immettere nel mercato delle due piazze quantità tali del prodotto in questione in modo da realizzare un prezzo medio di vendita nel complesso superiore a quello realizzato da A' ; ciò che a parità di costo unitario di produzione di A' e A significa un maggiore profitto unitario nella vendita per A .

Dal punto di vista formale il problema rientra nel nostro teorema se prescindiamo dalla legge della domanda, più precisamente è

analogo al caso degli stipendi degli insegnanti elementari e di scuola media ; la presenza, però, della legge della domanda introduce ulteriori vincoli oltre alla (2) del teorema, riducendo ulteriormente le possibilità di verificarsi del caso in questione che per comodità chiameremo « caso concorrenziale ».

Infatti, per la concorrenza che A vuole fare nei confronti di A' si dovrà avere :

$$\begin{aligned} a' &= f_1(\lambda' Q') > a = f_1(\lambda Q) ; \\ b' &= f_2[(1 - \lambda') Q'] > b = f_2[(1 - \lambda) Q] \end{aligned} \quad (3)$$

e, per quanto precede, dopo aver convenuto che $a' > b'$ si dovrà avere anche :

$$a > b' \quad (3)'$$

altrimenti non si potrà mai avere $m > m'$ (v. nota (3)) ; inoltre, per la decrescenza di $f_1(x)$ e di $f_2(x)$ dovrà aversi :

$$Q' \lambda' < Q \lambda < (1 - \lambda') Q' < (1 - \lambda) Q$$

in tal caso la condizione (2) del teorema diviene :

$$\left\{ \begin{aligned} 0 &< \frac{f_1(\lambda' Q') - f_2[(1 - \lambda') Q']}{f_1(\lambda Q) - f_2[(1 - \lambda) Q]} \lambda' + \frac{f_2[(1 - \lambda') Q'] - f_2[(1 - \lambda) Q]}{f_1(\lambda Q) - f_2[(1 - \lambda) Q]} < \lambda < 1 \\ Q' \lambda' &< Q \lambda < (1 - \lambda') Q' < (1 - \lambda) Q \end{aligned} \right. \quad (4)$$

che definisce così il « caso concorrenziale ».

Il verificarsi delle (4) o meno dipende da $f_1(x)$ e $f_2(x)$; infatti è facile vedere che il « caso concorrenziale » per certi tipi di legge di domanda è possibile, mentre per certi altri tipi è impossibile.

Se, ad es., $f_1(x) = k_1 x^{-1}$; $f_2(x) = k_2 x^{-1}$ (k_1, k_2 costanti positive) cioè se trattasi di iperboli equilateri, si vede subito che

il « caso concorrenziale » è impossibile, infatti in tal caso si ha :

$$m' = \frac{\lambda' Q' f_1(\lambda' Q') + (1 - \lambda') Q' f_2[(1 - \lambda') Q']}{Q'} = \frac{k_1 + k_2}{Q'}$$

ed analogamente :

$$m = \frac{\lambda Q f_1(\lambda Q) + (1 - \lambda) Q f_2[(1 - \lambda) Q]}{Q} = \frac{k_1 + k_2}{Q} = m' \frac{Q'}{Q}$$

ora, siccome nelle (4) è implicito che $Q > Q'$ ⁽⁴⁾ si ha che comunque si manovrano le quantità rimane $m' > m$.

D'altra parte, che il « caso concorrenziale » sia possibile per certi altri tipi di $f_1(x)$ e di $f_2(x)$ è subito visto mediante un semplice esempio numerico :

$$a' = 100 > a = 70 > b' = 50 > b = 49$$

$$\lambda' Q' = 1 < \lambda Q = 12 < (1 - \lambda') Q' = 99 < (1 - \lambda) Q = 108$$

$$\lambda' = 0,01 ; \quad \lambda = 0,10$$

$$Q' = 100 ; \quad Q = 120$$

$$\frac{a' - b'}{a - b} \lambda' + \frac{b' - a}{a - b} = 0,0714 < \lambda = 0,10.$$

$$m = 70 \times 0,10 + 49 \times 0,90 = 51,10 >$$

$$> m' = 100 \times 0,01 + 50 \times 0,99 = 50,50$$

Concludendo : il « caso concorrenziale » può aversi o meno a seconda della forma della legge della domanda.

5. Altra questione che presenta analogie con la precedente, sebbene formalmente diversa, è la seguente tratta da uno inte-

(4) Infatti da $\lambda' Q' < \lambda Q$ si nota che se fosse $Q' \geq Q$ si avrebbe $Q' - \lambda' Q' > Q - \lambda Q$ contro la disuguaglianza che figura nella (4) :

$$Q' - \lambda' Q' < (1 - \lambda) Q .$$

ressante studio di C. GINI « *Sulla durata del periodo pre-ovulare e della gestazione dei due sessi* » (« Clinica Ostetrica » n. 3, giugno 1960) nel quale figurano i seguenti dati :

Paese	Ricercatori	Data	Numero delle nascite			Durata media delle gestaz. in giorni		Differenza tra i due sessi in giorni $X_m - X_f = \Delta$
			maschi n_m	femmine n_f	totale n	maschi X_m	femmine X_f	
1. Londra	Karn, Penrose . .	1935-46	340	274	614	259,735	259,945	0,210
2. Londra	Karn, Penrose . .	1935-46	6.697	6.419	13.116	281,288	282,008	0,720
	in complesso . . .		7.037	6.693	13.730	280,246	281,105	0.859
1. Chicago	Yi-Yung Hsia . .	1955	73	50	123	275,51	276,08	0,57
2. Chicago	Yi-Yung Hsia . .	1955	71	73	144	279,05	279,19	0,14
	in complesso . . .		144	123	267	277,226	277,949	0,714

In questo prospetto si nota, considerando, ad esempio, le prime due inchieste di Karn e Penrose :

$$X_{f_1} - X_{m_1} = \Delta_1 = 0,210 ; \quad X_{f_2} - X_{m_2} = \Delta_2 = 0,720 ;$$

$$X_f - X_m = \Delta = 0,859 > \max (\Delta_1, \Delta_2)$$

cioè la differenza fra le medie nel complesso supera ambedue le differenze delle medie nei gruppi parziali ; ora si ha :

$$X_f = \lambda X_{f_1} + (1 - \lambda) X_{f_2} ; \quad X_m = \lambda' X_{m_1} + (1 - \lambda') X_{m_2} ;$$

$$(0 < \lambda, \lambda' < 1)$$

ed anche, tenendo presente che :

$$X_{f_1} = \Delta_1 + X_{m_1} ; \quad X_{f_2} = \Delta_2 + X_{m_2}$$

si ha :

$$\begin{aligned} \Delta = X_f - X_m &= \lambda X_{f_1} + (1 - \lambda) X_{f_2} - \lambda' X_{m_1} - (1 - \lambda') X_{m_2} = \\ &= \lambda \Delta_1 + (1 - \lambda) \Delta_2 + (\lambda - \lambda') (X_{m_1} - X_{m_2}) = \\ &= \underset{\lambda, 1-\lambda}{\text{med}} (\Delta_1, \Delta_2) + (\lambda - \lambda') (X_{m_1} - X_{m_2}) . \end{aligned}$$

ora, è chiaro che affinché sia :

$$\Delta \geq \max (\Delta_1, \Delta_2)$$

occorre e basta che sia :

$$(\lambda - \lambda') (X_{m_1} - X_{m_2}) \geq \max (\Delta_1, \Delta_2) - \underset{\lambda, 1-\lambda}{\text{med}} (\Delta_1, \Delta_2)$$

ossia si può affermare che affinché si abbia :

$$\min (\Delta_1, \Delta_2) \leq \Delta \leq \max (\Delta_1, \Delta_2)$$

occorre e basta che (come è facile riscontrare) :

$$\begin{aligned} \min (\Delta_2, \Delta_2) - \underset{\lambda, 1-\lambda}{\text{med}} (\Delta_1, \Delta_2) &\leq (\lambda - \lambda') (X_{m_1} - X_{m_2}) \leq \\ &\leq \max (\Delta_1, \Delta_2) - \underset{\lambda, 1-\lambda}{\text{med}} (\Delta_1, \Delta_2) \end{aligned}$$

oppure :

$$\begin{aligned} \min (\Delta_1, \Delta_2) - \underset{\lambda', 1-\lambda'}{\text{med}} (\Delta_1, \Delta_2) &\leq (\lambda - \lambda') (X_{f_1} - X_{f_2}) \leq \\ &\leq \max (\Delta_1, \Delta_2) - \underset{\lambda', 1-\lambda'}{\text{med}} (\Delta_1, \Delta_2) \end{aligned}$$

Le espressioni :

$$(\lambda - \lambda') (X_{m_1} - X_{m_2}) ; \quad (\lambda - \lambda') (X_{f_1} - X_{f_2})$$

possono interpretarsi come *interazioni* della variazione del peso da λ a λ' (tra maschi e femmine) e delle medie tra i due gruppi per lo stesso sesso.

È, in sostanza, un'adeguata limitatezza in valore assoluto delle suddette *interazioni* che consente a Δ di cadere entro l'intervallo :

$$\min (\Delta_1, \Delta_2) \text{ — } \max (\Delta_1, \Delta_2)$$

Ora, nei casi numerici in questione è :

$$1) \quad \lambda = \frac{274}{6693} = 0,0409; \quad \lambda' = \frac{340}{7037} = 0,0483$$

$$X_{f_1} - X_{f_2} = -22,063; \quad X_{m_1} - X_{m_2} = -21,553; \quad \max(\Delta_1, \Delta_2) = 0,720$$

$$\text{med}_{\lambda, 1-\lambda}(\Delta_1, \Delta_2) = 0,6991; \quad \text{med}_{\lambda', 1-\lambda'}(\Delta_1, \Delta_2) = 0,6954$$

$$(\lambda - \lambda')(X_{m_1} - X_{m_2}) = 0,1595 > \max(\Delta_1, \Delta_2) - \text{med}_{\lambda, 1-\lambda}(\Delta_1, \Delta_2) = 0,0209$$

$$(\lambda - \lambda')(X_{f_1} - X_{f_2}) = 0,1633 > \max(\Delta_1, \Delta_2) - \text{med}_{\lambda', 1-\lambda'}(\Delta_1, \Delta_2) = 0,0246$$

$$2) \quad \lambda = \frac{50}{123} = 0,4065; \quad \lambda' = \frac{73}{174} = 0,5069$$

$$X_{f_1} - X_{f_2} = -3,11; \quad X_{m_1} - X_{m_2} = -3,54; \quad \max(\Delta_1, \Delta_2) = 0,57$$

$$\text{med}_{\lambda, 1-\lambda}(\Delta_1, \Delta_1) = 0,31479; \quad \text{med}_{\lambda', 1-\lambda'}(\Delta_1, \Delta_2) = 0,34537$$

$$(\lambda - \lambda')(X_{m_1} - X_{m_2}) = 0,3554 > \max(\Delta_1, \Delta_2) - \text{med}_{\lambda, 1-\lambda}(\Delta_1, \Delta_2) = 0,25521$$

$$(\lambda - \lambda')(X_{f_1} - X_{f_2}) = 0,3122 > \max(\Delta_1, \Delta_2) - \text{med}_{\lambda', 1-\lambda'}(\Delta_1, \Delta_2) = 0,22463$$

SUMMARY

The A. takes the hint from numerical concrete situations about the wages distribution in Italy of particular categories of teachers, to establish the *necessary* and *sufficient conditions* in order that a weighed average be greatest than another.

More precisely, the proposition proved is the following :

THEOREM : *Let $m = \lambda a + (1 - \lambda) b$ and $m' = \lambda' a' + (1 - \lambda') b'$ with $a, b, a', b', \lambda, \lambda'$, reals of which $a > b$ and $0 \leq \lambda, \lambda' \leq 1$, then in order that it be :*

$$m \geq m'$$

it is necessary and sufficient that it be :

$$\max (0, c \lambda' + d) \leq \lambda \leq 1$$

where

$$c = \frac{a' - b'}{a - b}; \quad d = \frac{b' - b}{a - b}.$$

Besides, it is examined the interesting case of $a = f_1(\lambda Q)$ and $b = f_2[(1 - \lambda) Q]$ where Q is a constant, f_1 and f_2 two monothonic functions; similary for $a' = f_1(\lambda' Q')$; $b' = f_2[(1 - \lambda') Q']$.

In particular, it is examined the case in which a, b and a', b' are prices and f_1, f_2 are the laws of demand of the same product in two different places or situations.

As consequence of the above, the Author shows as the « concurrence case » (defined by :

$$\left\{ \begin{array}{l} 0 < \frac{f_1(\lambda' Q') - f_2[(1 - \lambda') Q']}{f_1(\lambda Q) - f_2[(1 - \lambda) Q]} \lambda' + \frac{f_2[(1 - \lambda') Q'] - f_2[(1 - \lambda) Q]}{f_1(\lambda Q) - f_2[(1 - \lambda) Q]} < \lambda < 1 \\ Q' \lambda' < Q \lambda < (1 - \lambda') Q' < (1 - \lambda) Q; \\ f_1(\lambda' Q') > f_1(\lambda Q) > f_2[(1 - \lambda') Q'] > f_2[(1 - \lambda) Q]. \end{array} \right.$$

where Q is the output to be sold and λ the fracional part of Q referred to the firm A and Q', λ' are the analogue magnitudes of the firm A' for which $m \geq m'$ although it be $a < a'$; $b < b'$, it can verified only for particular laws of demand. For instance, in this note it is showed that if f_1 and f_2 are equilateral hyperbolas the « concurrence case » is not possible, while it is possible in the case of other laws of demand.

CORRADO GINI

**Statistical Methods with special reference
to Agriculture (1)**

CHAPTER I

GENERAL CONSIDERATIONS

I. I. ORIGIN AND DEVELOPMENT OF STATISTICS

I. I. I. Administrative preparation.

In order properly to understand the present concept of Statistics, it is opportune to examine how this concept has been developed through the ages.

Records of statistical data or surveys go back as far as the historical and archaeological knowledge that we possess of the past. But we may affirm with certainty that there were statistical data and operations even in earlier periods, though in the simplest form, ever since human communities began to exist.

Upper Paleolithic hunters apparently kept count of their season's kills on notched sticks. These may be regarded as the first statistical documents.

In Sardinia there are ancient monuments called *Nuraghi*, on whose walls are sometimes to be found rough signs which have been interpreted as an enumeration of the products of agriculture or of hunting, traced in a period for which there is no evidence as yet of the existence of writing.

(1) From lectures given at the International Center for Training in Agricultural Economics and Statistics at Rome, January-May, 1956 in cooperation with the Government of Italy, the University of Rome, the Food and Agriculture Organization of the United Nations and the International Statistical Institute. Two Chapters of said lectures have been already published in the two last issues of this journal. See: *Logic in Statistics*, «Metron», Vol. XIX, N. 1-2; *Mathematics in Statistics*, «Metron», vol. XIX, N. 3-4.

Little by little, with the progress of social groups towards the shaping of the State, the enumerations and the collections of statistical data became not only more frequent, but also less uncertain and less imperfect.

The old Egyptian monuments prove that the Egyptian people were following the state and movements of population by the taking of censuses and by means of regular collections of data.

Statistical documents have been found in the library of Assurbanipal, King of Assyria, and we know of censuses carried out by other Eastern peoples, such as for the Jewish people the census in the desert made by Moses, the census of King David in his kingdom and that made on the return from slavery in Babylon.

In a book written about 559 B.C. Confucius relates that the Chinese king, Yao, in the year 2218 B.C. ordered agricultural, industrial and commercial statistics to be collected in his land.

The distribution of land, military service and the organization of education of young generations certainly needed regular statistical surveys in Greece.

In Roman times the collection of statistical data was highly developed. The census did not only ascertain the number of citizens but also their wealth and constituted the basis of the state organization. Births and deaths used to be declared in connection with religious offerings.

After the fall of the Roman Empire, the vestiges of the collection of statistical data disappeared from Europe, while they showed an increasing development among the Arabs.

In Europe they flourished again with the restoration of the State organization carried out by Charlemagne.

In England, William the Conqueror (eleventh century) had a kind of cadastral survey of real estate drawn up, called the *Liber Censualis* or Domesday Book, which constitutes a statistical document of great importance.

At the end of the thirteenth century, the flourishing Italian republics carried out many administrative and political acts having a statistical character, such as the enumeration of citizens fit for military service and the reports to their Government by the Venetian Ambassadors. In Spain, too, periodical collections of data relating to population have been carried out since the fourteenth century.

The Church, acknowledging the importance of statistical returns, established compulsory registration of marriages, baptisms, burials after the Council of Trent (16th century).

I. 1.2. *Literary preparation.*

What we have called the administrative preparation of Statistics developed side by side with another kind of preparation that might be called literary, in so far as it is contained in the works of historians, geographers, politicians, agrarians, etc.

In a dialogue between Socrates and Glaucones, for instance, evidence is adduced on the necessity for statesmen to develop statistical investigations. In Herodotus' History, as well as in Aristotle's Republic and the works of his followers, descriptions can be found, though often in qualitative form, of those phenomena relating to state organization which later become the subject of statistical returns.

Pliny and Phlegethon, Ulpian as well as the Greek and Roman historians relate many statistical data, concerning for instance, the number of centenarians, the average life to be expected at each age, the number of soldiers, the losses in battles, in epidemics, etc.

In the Middle Ages reports of that kind are no longer found in Europe but developed in the works of Arabian writers.

They appear again with the restoration of European States, in Villani's *Cronache*, in Brunetto Latini's *Tesoro* and especially in a booklet in which Marion Sanudo the Old advocated the reconquest of the Holy Land from the Infidels. The mass of data becomes larger and is directed to a more general scope in the works of Sebastian Münster and John de Laet, as well as in Francesco Sansovino's book issued in 1551 with the title, *Del governo et amministrazione di diversi Regni et Repubbliche così antiche come moderne*, and in Giovanni Botero's work, *Le relationi universali*, published in 1582. All these authors, whose works have been widely known, can be regarded as the forerunners of scientific Statistics.

Actually, the last-mentioned writers give not only facts and data relating to population, production, etc., but consider the description of States as a branch of study in itself, independent of other branches.

I. 1.3. *The origin of scientific Statistics - Hermann Conring and Descriptive Statistics.*

The general acceptance of the description of States as an independent branch of study followed in Germany in the middle of the seventeenth century with the work of Vitus von Seckendorff (1626-1692) and especially with that of Hermann Conring (1606-1681), who is considered the founder of this tendency in Statistics.

Vitus von Seckendorff, opposing, in a volume published in 1656, the idealistic literature of his time, declared his intention of directing the new study to practical purposes in the life of States, a life which he describes from the threefold aspect of territory, government and finance.

Hermann Conring, a professor of philosophy, medicine and politics at Helmstädt University, greatly improved this new tendency. In 1661 he started a new course of teaching in political science, in which the notions and data relating to the life of States were elaborated and systematized in the scholastic divisions. The scope of the course was directed to the description of the notable things of the State.

Conring had a great number of followers. One of them, Schmeitzel, had been the teacher, at the University of Jena, of Gottfried Achenwall (1719-1772), whose fame became so great that it obscured Conring's. His teaching at the University of Göttingen took the name of *School of Göttingen* and then of *Achenwall Statistics*. No special knowledge had been added by him to the teaching of the founder, Conring, but he had the great merit of having established the position of the new discipline in academic and official circles, defending it from the objections raised by the professors and politicians, and he gave it the name of *Statistics*.

Conring has been called the Columbus of the new science ; Achenwall may be said to have been its Amerigo Vespucci, even though the word "statistics" had been used long before (1589) in Italy by Ghilini.

A crisis in the new science was caused by the introduction of the tables made by the Danish historian, Anchersen (1799-1865), in his *Descriptio statum cultiorum in tabulis*. This work

contained synoptic tables with a rather limited number of figures, but, limited as they were, these tables represented a new method, which was developed as more data became available.

The representative of traditional Statistics, the so-called *Universitarian Statistics* strongly opposed such a method, because, attached to Conring's ideas, they still considered Statistics as a *description* of the notable things of the State, which at the time could not be expressed in figures except for some isolated topics.

The method of tables, however, went ahead through the works of Crome (1753-1833), an official defender of it, and, before, due to the authority of von Schlözer (1735-1809), the father of German journalism, whose name should be remembered beside those of Conring and Achenwall, as that of a meritorious organizer of Statistics, even from the point of view of official collections.

Crome also introduced graphical representation, and Schlözer opened the way to historical Statistics, while another author, Büsching, a geographer, introduced the comparative method.

Universitarian Statistics now had a name and a system, as well as its technical means and public opinion in its favour. Its object continued to be the description of important things in the State, but it was now a *quantitative* description, so that henceforth it approached the modern conception of Statistics.

However, in order to reach the level of a science, the new study had to go beyond the pre-scientific level of the description of facts and pass to the search for laws, thus fulfilling the aim of science.

The thesis of the economist J.B. Say (1767-1832), who claimed that Statistics, as a pure descriptive study, could not be qualified as a science, provoked a strong reaction, especially in Italy where the Achenwallian conception was applied to several sciences with a modern breadth of view by M. Gioia (1767-1829), G. Romagnosi (1761-1835) and L. di S. Cagnazzi (1764-1852), not only describing but also interpreting the facts and establishing their regularities. But the elevation of Statistics to a discipline primarily investigative in character was due to another current of thought, developed in England at the same time as the universitarian Statistics in Germany, and namely that of the *Political Arithmeticians*.

I. 1.4. *Biometricians, Political Arithmeticians and Investigative Statistics.*

The work of John Graunt (1620-1764), presented in 1661 to the London Royal Society, dealt with *Natural and political observations upon the bills of mortality, chiefly with reference to the government, religion, trade, growth, air, diseases etc., of the city of London.*

Data on baptisms and burials were extracted from the registers of the churches, and published week by week. Though rough and incomplete, these data were consulted, for information concerning contagion, by rich people who wanted to leave the city in good time, as well as by traders, who made forecasts for their business for a future season. Graunt was able to elaborate from this material a series of scientific uniformities concerning the equilibrium of the sexes in the population, the excess of male over female births, the frequency of insanity and suicides, etc. He also outlined, using scanty information, a table of mortality and calculated the population of London.

With Graunt's work the empiric inductive method of natural science was transferred to social science and gave the science of population as its first result.

Single investigations on population problems had been carried out in Italy before Graunt's research. In 1570 the mathematician Cardano, had prepared a theoretical table of mortality based on inductions even more audacious than Graunt's, and the excess of male over female births has been known in Florence since 1300. But there was not yet any organic work of investigative Statistics in this field.

Such a work had, however, been published, in his treatise *De statica medicina*, in another field by Santorio di Capodistria (1561-1636), who introduced into the study of man the quantitative methods of natural science later, adopted by Graunt for the study of population; he also measured the pulsations of the heart as well as the basal metabolism, studying their dependence on weather, age, food, health, etc.

Santorio's methods were widely echoed in the medical world, especially in England, where a school called the *Iatromathematical School* started to work on quantitative biological research.

Santorio was a pioneer, in Europe, of what today is known as *Biometric* and *Sanitary Statistics*, but in India *Somatometrics* was already being studied systematically and applied in the ninth century B.C.

After Graunt's work on London, William Petty (1623-1687), a man of very wide culture and of high scientific and social position, accomplished a similar task for Dublin, together with various essays on the new science, which he called *Political Arithmetic*, and of which he was considered the head. In Petty's work one should mention the calculation of the period of time in which a population could be doubled, a subject that was dealt with by Malthus a century later.

Later, with Gregory King (1648-1712) and Charles Davenant (1657-1714), Statistics dealt not only with population but also with financial problems, while the astronomer Halley (1656-1742) introduced and applied a method for the calculation of the tables of mortality.

Of *Political Arithmetic* Davenant gave the following definition: "The art of reasoning by figures upon things relating to Government".

Such a definition not only shows the similarity of purpose (the public interest) and means (figures), but also the difference (reasoning instead of description) when compared with the quantitative description of the notable things of the State, according to the Achenvallian current of thought. Therefore Political Arithmetic was called *Investigative Statistics* as opposed to the German Universitarian Statistics known as *Descriptive Statistics*.

In 1671, the Dutch mathematician and outstanding politician, Jan de Witt, worked out a report, presented to the legislative bodies of Holland, on life annuities based on a law of survival that he had established. At the same time two other tables of mortality were published by van Dael and Hudde, while prominent people such as Leibniz and Vauban advocated, although without success, the introduction of systematic statistical collections of data. Inquiries into the amount and movement of population were declared compulsory in Sweden in 1748.

In France in the eighteenth century statistical research was followed with interest by people such as Necker, Lagrange, Laplace and Lavoisier. France and Holland, the typical countries

of life annuities, improved the accuracy of mortality tables. Meanwhile the great Swiss mathematician, Euler, showed the correction to be introduced into Halley's method for taking account of the excess of births over deaths in compiling tables of mortality, while in England Th. Simpson and R. Price tried to take account of migratory movements.

I. 1.5. The Demographic current.

Demographic Statistics developed quickly and reached maturity in a work by J.P. Süßmilch (1707-1767), *The divine order in the variations of the human kind, demonstrated through their births, deaths and reproduction*. Issued in 1741 but generally known in its posthumous edition of 1775, it deals with the laws of population movements, laws that the author seeks passing by means of mass observation from the variable to the uniform, that is to say using a procedure which is considered, even today, as typical of Statistics.

Süßmilch's work is a genuine treatise on population and gives evidence on every aspect of the problem: numerical economic, sanitary, moral and theological. It is thus no wonder that his work obscured that of Graunt, the founder of the new current, nor that Investigative Statistics was called the *School of Süßmilch*, in the same way that descriptive statistics had been known as the *School of Achenwall*.

Süßmilch did not, however, give the name to the discipline he had systematized, which received in the following century the name of *Demography* through the initiative of Achille Guillard.

I. 1.6. Calculus of Probability and encyclopaedic-mathematical tendency.

Another tendency issued from the school of Political Arithmeticians. It can be called mathematical because in its investigations, following the tendency of the Political Arithmeticians, it made considerable use of mathematics and, as it extended investigation to all fields of science, it can be called encyclopaedic.

The powerful tool which permitted the flourishing of this tendency was the Calculus of Probability.

Its roots are in the games of chance, games as old as human intelligence. Archaeological findings suggest that 40.000 years ago our ancestors were gambling with ankle bones of sheep, goats and deer.

Between that epoch and the third millenium B.C. games of chance were organized. Frequently ankle bones and dice, later lotteries and cards, finally roulettes and other mechanisms were their instruments.

It is very likely that everyone who invented or practised such games was interested in knowing the chances of the different possible issues, but notices which attest a wide interest of scientific people in such problems carried us on to the Italian Renaissance.

It is not without reason that in this tendency of Statistics, as well as in the descriptive tendency and the investigative tendency, we find Italian leadership, as Italy was, during this period, at the head of civilization, and the development of Statistics is strictly bound up with cultural progress.

Luca Pacioli (1445-1514), Tartaglia (1499-1557), Cardano (1501-1576), Peverone (1509-1559) and Galileo (1564-1642) have left us the results of their pioneer studies on the matter, sometimes condensed in special treatises, as in the case of Galileo and especially of Cardano. Cardano especially was not concerned only with the solution of special problems arising from the dice gambles, but intended to give general criteria for the calculation of chances.

These were certainly not the first treatises on games of chance. The Emperor Claudius, a passionate dice gambler, had already, in the first century A.D., written a treatise on dice, but it did not reach us and it is not supposed to have been mathematical in character.

It is very probable on the contrary, that mathematical problems in games of chance had been treated, before the Italians, by Arab scholars who were ahead of the Europeans in mathematics during the Dark Ages, so that it is quite credible that Al-Ghazali solved a problem similar to those that were treated some centuries afterwards in Italy and France.

As a matter of fact, if Italy was the cradle of the Calculus of Probability in the fifteenth and sixteenth centuries, its study

was resumed in the seventeenth century in France by Pascal, (1623-1662) and Fermat (1601-1665). Christian Huygens (1629-1695), a young Dutch scholar destined to become prominent in the mathematical and physical sciences, on a visit to Paris, contacted some French mathematicians and, returning home, in 1657 wrote the treatise *De ratiocinio in ludo aleae* in which the mathematical problems of games of chance were organically expounded.

So far, however, the researches were in the field of pure mathematics, now called Combinatory Calculus. They could be applied only to games of chance in so far as these are "honest", i.e. constructed on the basis of the equal chances of all the equally conceivable results.

The concept of probability was not yet familiar to the mathematicians, and for a long time afterwards it was kept distinct from that of chance.

Probability was called the faith in the foundation of a judgement concerning a certain happening. To extend the Calculus of Chances to the probabilities was the great achievement of James Bernoulli (1654-1705), who consequently is rightly considered the father of the theory of probability. This passage was made by measuring the probability of a judgement either *a priori*, through the number of chances favorable to the happening to which the judgement is referred, divided by all the possible chances, or *a posteriori*, through the number of cases in which the happening occurred divided by all the cases in which it could have occurred.

Eventually the probability of the judgment concerning a happening was transferred to the happening itself, and it is in that sense that we now speak of the probability as synonymous with chance.

James Bernoulli not only, in this manner, laid the foundations of the theory of probabilities, but with the celebrated theorem which has taken his name constructed upon them the Calculus of Probability.

The transfer to the probabilities of the theory of chances as well as the application to them of the theorem of Bernoulli were not exempt from far-reaching hypotheses, which from the very beginning had provoked objections from Leibniz, to whom Bernoulli had communicated the results of his researches before-

hand, but which were often forgotten or neglected in the enormous development that the Calculus of Probability assumed, due to the contributions of De Moivre, Daniel Bernoulli, Euler, d'Alembert, Condorcet, Lagrange, Laplace, Poisson, Gauss, Bienaymé, De Morgan. All these were famous mathematicians who had the great merit of endowing Statistics with a powerful means of investigation, but in the applications to everyday life many of them and the generality of their followers went too far.

The Calculus of Probability teaches how to deduce, from the probability of elementary phenomena, the probability of more complex ones which are the result of combinations of the former.

Those investigating this Calculus were carried away with enthusiasm for their discovery and believed that they could forecast any kind of happening by means of it, as soon as they knew the probabilities of the elementary phenomena of which the happening was the out come. Therefore they applied the Calculus of Probability to all kinds of phenomena — judicial, electoral, meteorological, physical — and not only to the social and political phenomena examined by the school of Political Arithmeticians.

Not taking into account the hypotheses on which the theory and the Calculus of Probability were based, they sometimes arrived at absurd results which made people speak of "the scandal of the geometricians" and did a good deal of harm to the new branch of science.

A more cautious attitude was subsequently taken and, on the other hand, some of the above-mentioned mathematicians, such as Gauss and Bienaymé, had been more rigorous in their conclusions than their predecessors. It was eventually recognized that the Calculus of Probability could be rightly applied to concrete phenomena only if the hypotheses on which the measurement of probabilities and their calculus stand, prove true in reality.

Two outstanding authors took this attitude in their applications of the Calculus of Probability to statistical phenomena, in conformity with the encyclopaedic-mathematical tendency. They are Adolphe Quetelet and Augustin Cournot.

Quetelet (1796-1874), a Belgian astronomer and statistician, is often considered the founder of modern statistics. His main

work, which appeared in 1838, is entitled, *Sur l'homme et le développement de ses facultés, ou essai de physique sociale*. In this work numerical research is based on the Calculus of Probability; the author establishes the basis of statistical methodology and uses the method, now scientific, for his study of population and of the physical and moral qualities of man, thus establishing the bases of his own system of science and philosophy.

He subsequently published *Le système social et les lois qui le régissent*; in 1869 made a new edition of the *Physique sociale* and in the following year produced his last work, entitled *Anthropométrie*.

His activity was also outstanding in practical problems; he was director of the astronomic observatory of Brussels and promoted the International Congresses of Statistics.

Antoine Augustin Cournot (1801-1879), a French mathematician, economist and philosopher, in his work: *Recherches sur les principes mathématiques de la théorie des richesses* formulated some of the most important problems of mathematical economy. Concerning the theories of the Calculus of Probability and their application to Statistics, he contributed one of the most discerning treatises that have ever been made, *Exposition de la théorie des chances et des probabilités* (1843).

I. 1.7. *Gustav Rümelin and the eclectic period.*

Investigative and Descriptive Statistics had been developed independently; when they came into contact in Germany, it was found that there was much to be discussed. The economist Knies made it clear that they were two separate branches that could not be confused; the one is the Statistics that aims at recognizing the regular nature of human society; the other is the Statistics that aims principally at describing the situation of the more important things in a State. Gustav Rümelin changed this division into a threefold division, when he divided investigative statistics into a technical methodological part and an applied part.

During this period, which we can call eclectic, Statistics is considered from three distinct points of view: there is a Methodological Statistics, a general method of investigation adapted for certain phenomena (this concept was predominant in the

encyclopaedic-mathematical tendency and was clearly expressed by Cournot); there is a Social Statistics investigating the laws of society from the quantitative point of view (the conception which was accepted by the demographic tendency, as in Süssmilch); and there is a Descriptive Statistics which studies the important facts of the State, as we have seen in the teaching of Conring and Achenwall, from the quantitative point of view.

I. 1.8. The modern concept of Statistics and its recent characteristics.

In time the threefold division of Statistics mentioned above became insufficient, numerous statistical investigations being made in every field. Thus, for example, statistical investigations on the human body made by Quetelet (who called them Anthropometrics), as well as other investigations on living organisms made by Galton and Galton's followers (especially K. Pearson) could not be included in the threefold division. These studies, representing quantitative biology were united under the name of Biometrics.

As soon as statistical investigations and applications were extended to new fields and phenomena, there were new divisions of Statistics. Thus, for instance, from Quetelet's and Guerry's researches arose "Criminal Statistics": from von Oettingen's research on moral phenomena, "Moral Statistics": from research in psychology, "Psychometrics": from that on physical phenomena, "Mechanical Statistics", while statistical method applied to sociology produced "Sociometrics" and applied to history, "Historiometrics".

Meanwhile quantitative statistical methods were applied in ever increasing measure to Genetics, Eugenics, Botany, etc.

Inductive investigation had begun in the field of economic (Pareto) and political phenomena and it was realized that the description of the important things in a State was only the first stage of a political and economic science having the same investigative character that Demography had already assumed. Economic investigations carried out by the use of quantitative methods have acquired the name of "Econometrics".

In this period many general treatises have been published, as well as monographs on methodological statistics. Particular mention should be made of the works of von Mayr, Benini, Yule, Charlier, Westergaard, Bowley, Gini, Davenport, Czuber, March,

Julin, R.A. Fisher, Niceforo, Livi, Boldrini, Bachi, Snedecor, Vinci, Kendall, Vianelli, etc.

According to the modern concept, there is therefore, on the one hand, "Methodological Statistics" and on the other "Applied Statistics", which in turn is divided into branches such as Demography, Biometrics, Anthropometrics, Econometrics, Psychometrics, Sociometrics, etc.

Descriptive Statistics must, therefore, be considered as the first stage of these applications; it is, however, a stage of paramount importance, to which the great majority of statisticians are dedicated, collecting data, elaborating and publishing them.

The greater part of official statistical work is descriptive in character, although in recent years some scientific tasks needing vast cooperation and being of national or world interest, such as the estimation of national income, have been taken over by State or international organisations. But also private firms and corporative or cooperative bodies, in the regulation of their own economic activities or the putting of more detailed or fresher data than official ones at the disposal of the public, make important statistical surveys.

Another characteristic of the most recent development of Statistics is the importance assumed by mathematical methods, the treatment of which is only in a minority of cases accompanied by applications to concrete phenomena. This procedure gives rise to serious reservations. It is not possible to judge the usefulness of a method except from its applications. On the other hand, statistical methods, particularly those based on the Calculus of Probability, imply hypotheses for which a verification is indispensable, otherwise, the doubt is authorized that the conclusions obtained are not founded, especially when they are not in agreement with those that experience suggests to the specialists. It would not be the first time that the upper hand taken by mathematicians in science has had the effect of neglecting the substantial aspect of the researches in comparison with the formal one, sometimes leading to extravagant conclusions. This was the case at the time of Pythagoras as well as at the time of Laplace. And some of the more responsible statisticians have the feeling that, if a reaction does not intervene, the indiscriminate applications of the refined mathematical methods in

the statistical field are leading once more to a "scandal of the geometricians" of an extent much greater than the past one.

In the course of these lectures I shall often call the attention of the audience to the hypotheses that the statistical methods imply, and to the frequent misuse of such methods through not duly taking account of the hypothesis involved.

I.2 ELEMENTARY NOTIONS

I. 2.1 *Concept of Statistics*

In order to understand what Statistics is, the needs from which it arises and which must be satisfied have to be taken into consideration.

We can realize that there are phenomena which cannot be perceived through a single observation, but which require a synthesis of observations of other, more simple, phenomena. The former are called *mass* or *collective phenomena* and the latter *single* or *individual phenomena*. For instance, the amount of population represents a collective phenomenon deriving from the observation of individual persons, who therefore represent the single phenomena; the coefficients of natality, nuptiality and mortality also represent collective phenomena, while the single births, marriages or deaths are the individual phenomena from which they derive. Thus the necessity arises for a technique especially adapted to the quantitative investigation of collective phenomena; Methodological Statistics fulfils this want.

Statisticians in the past have been interested in the problem of knowing if Statistics should be considered as a science or simply as a method.

We believe that, rather than a method, it should be defined as a technique.

We can therefore define Methodological Statistics as "a special technique suitable for the quantitative investigation of mass or collective phenomena; those phenomena (physical, biological and social) being considered mass or collective phenomena whose measurement requires a collection of observations of other simpler phenomena, called single or individual phenomena".

In practice, however, it must be observed that, if Statistics may be considered as a pure technique, when it is taught together with the branch of science to which it is applied, it may be considered as a discipline in itself, when dealt with independently.

Yet a question arises which is important from the didactic point of view, and that is, whether it is preferable for Statistics, as a necessary technique for the study of various sciences, to be taught as propaedeutic to each of these, or to be taught as an independent discipline accompanied by examples of applications which can be taken from all sciences.

In the United States the first system, and in Europe the second prevailed for a long time. Experience shows that the European method gives the best results and, in fact, the European contribution to Methodological Statistics was higher than that which America had made in the past. This may be explained by the fact that applications in some branches may encourage the development of statistical methods, which may be beneficially applied to other branches. Moreover, the application of statistical methods requires a special faculty for realizing by intuition which are the most important elements influencing a given phenomenon, and on which investigation must consequently be concentrated, as it would be impossible to take all the elements into consideration. It is difficult for all the specialized scientists to possess such intuition, while people who possess it may more successfully carry out the application of statistical methods to a large group, if not to all, of the sciences to which Statistics can be applied.

As the latest development of the European system, special Schools of Statistics have been founded in several Italian universities since 1928 and a Faculty of Statistical, Demographical and Insurance Sciences was eventually set up in Rome in 1936.

After the second world war, special Schools of Statistics were founded in several other countries, such as those in North Carolina (U.S.A.), Paris, Sofia, Leningrad, Argentina, Madrid and Rio de Janeiro, while Departments of Statistics have been established in many universities both in Europe and America, as well as on the other continents. The European system of teaching has been adopted even in America with very fruitful results, and is becoming general.

I. 2.2. Field of application of Statistics

Phenomena to be dealt with statistically are to be found in every science and can be divided into three big groups :

a) Phenomena that cannot be observed through simple observation.

This may happen :

a.1) because the number of individual cases is so large that our minds become lost in numbering them ; for instance, the number of individuals composing the population of a State ;

a.2) because the cases appear at such intervals of time that our minds cannot make a connection between them ; for example, the number of centenarians who died during a certain period ;

a.3) because the individual phenomena of different kinds do not come under observation with the same frequency. For instance, the first travellers of India related that there were ten women to every man. It was established later that such proportion must be reduced to 107 to 100. The first travellers, being especially in touch with the upper classes, met people with many wives and did not observe that poor people could not afford to have more than one. Under these circumstances, cases of the phenomenon presenting different characteristics fell to the attention of observers with different frequency. Similar mistakes have been made by zoologists as to the proportion of the sexes among some species. They believed that certain kinds of animals had an enormous excess of males, while regular statistical collections proved that there is only a small excess of them, and sometimes an excess of females. The males being more visible and more active than the females, had more frequently come under the eye of the investigator ;

a.4) because phenomena of different kinds leave different impressions in our minds and are recorded in our memories in a different degree. Cases contrasting with our expectations impress our attention more than those which correspond to our wishes. For instance, Huarte de San Juan, a Spanish surgeon, wishing to give some idea of the proportion between the sexes at birth, wrote "Generally 6-7 females are born to one male." Statistical investigation ascertained later that, on the contrary,

in human birth there is a regular excess of males over females (approximately 6 per cent). Huarte's mistake, then, had a psychological basis in the common preference for the birth of a male. Contrasts also leave a deeper impressions than likenesses. This explains the current opinion concerning the existence of a tendency to marry between individual with different characters, as well as that of a contrast between professional success and academic career, while statistical investigations prove that there is a strong positive correlation between academic career and professional success, and that a tendency to marry exists between persons of similar character ;

b) Phenomena that require the application of statistical technique if they are to be observed, not only from the qualitative, but also from the quantitative point of view.

For instance, it is a common observation that young people marry young people, that the height of men is generally greater than that of women, that children resemble their parents, that good conditions of environment and food reduce mortality, but it is only through the use of statistical technique that we can transform these qualitative impressions into a quantitative measure of such phenomena.

c) Phenomena which may be observed, even quantitatively, without the help of statistical technique, but in which more or less serious errors can be made and have to be eliminated through the use of statistical technique.

Such are the phenomena that could be quantitatively investigated with a certain approximation even by one observation (e.g. the length of a table, the weight of a body, etc.). This single observation, however, cannot give us the exact measurement we may require. The proof is given by the fact that, by repeating an observation many times, we notice slight differences in the result, as the figures are influenced by casual mistakes. To obtain the value corresponding to reality or nearest to it, we need to have recourse to statistical technique.

Consequently we may say that the sphere of application of Statistics embraces all quantitative phenomena as well as those qualitative phenomena which can be reduced to quantitative phenomena. We may also say that all, or almost all, scientific

phenomena that need to be investigated exactly, ought to be investigated statistically.

In practice, however, the statistical technique, which is long and sometimes laborious and expensive, is more or less advisable according to the nature of the phenomenon investigated.

Sciences which require statistical technique may be divided into three groups corresponding to the groups of phenomena mentioned above.

A) Sciences relating to phenomena that, because of their nature, cannot be studied without Statistics. The typical science in this group is Demography.

B) Sciences that can be studied qualitatively but not quantitatively by means of common observation; such are Biology, Anthropology, Criminology, Eugenics, History, Sociology, Economics, Finance, etc. The subjects of these sciences may be dealt with without reference to statistical technique, but a serious study of some parts of them cannot be made without it.

C) Sciences requiring great precision in their results (Astronomy, Geodesy, Meteorology, Physics, etc.). Statistics constitutes a necessary complement of the studies in these sciences or, at any rate, of the studies of some of their aspects.

I. 2.3. Applications of Statistics to Agriculture

Agriculture is a field in which the applications of statistical techniques are particularly important.

These applications are as old as agriculture itself. We have already seen that one of the most primitive records of statistical collection of data is that preserved on the walls of the *Nuraghi* of Sardinia, possibly concerning agricultural products.

Crops are a collective phenomenon par excellence, the measurement of which cannot be satisfactorily effected by current observation. Their wide range makes any estimation difficult; their importance for human welfare creates an interest and raises expectations which may easily lead to biased impressions.

On the other hand, the yield of crops varies considerably from time to time and from place to place.

Moreover, agriculture extends over the whole world and, in view of the high degree of trade which takes place from one

country to another, the people of one country are interested in the agricultural production of other countries.

Special difficulties arise from the fact that in many countries the people engaged in agriculture are very backward from the cultural point of view and particularly incapable of dealing with figures.

Collection of data which cover all the cultivated land is therefore practically excluded and, in order to obtain reliable estimations, it is necessary in all countries to have recourse to partial returns; this implies the measurement of the influence of chance and, therefore, the application of most refined procedures of the Calculus of Probability.

Reliable estimations have, on the other hand, become necessary in recent times, in order that the countries which are associated in the United Nations may know in what direction and to what extent their common efforts for general progress have to be made.

For these reasons, special Centers directed to the diffusion of a knowledge of statistical methods and of their applications to Agriculture have been established by the Food and Agriculture Organization of the United Nations in many countries. The Training Center in Agricultural Economics and Statistics instituted in Rome and sponsored jointly by the Government of the Italian Republic, the University of Rome and the Food and Agricultural Organization of the United Nations has a special character, being of a higher level and recruiting students who already have a university degree and consequently some statistical knowledge.

This course, entitled *Statistical Methods with special reference to Agriculture*, has the aim of reviewing statistical methodology and analysis from a general point of view, illustrating them with examples relative to agriculture. The chapters on *Sampling Methods* and on *Correlation Methods*, as well as that on *Data collection, processing and publication* will be the subject of detailed treatment in special courses.

I. 2.4. Collective or mass phenomena

As we said above, when speaking of the concept of Statistics, collective or mass phenomena are those whose measurement requires a mass, or collection, of observations of other simpler

phenomena which are known as single or individual phenomena, and of which the collective or mass phenomena are composed.

In some treatises collective or mass phenomena are defined as phenomena which refer to a mass of single phenomena. This definition is inexact. For example, the percentage of males amongst births and the percentage of males amongst people marrying both refer to a mass of individual phenomena, but the first, not being ascertainable otherwise than statistically by means of a mass of observation, is a collective phenomenon, whereas the second does not require any observation, since it is known "a priori" that the percentage of males amongst people marrying is 50 per cent, and is therefore not a collective phenomenon.

I. 2.5. Statistical Unit and Statistical Datum

The result of an observation on one of the individual phenomena which make up the collective phenomenon is called a *statistical unit*. For example, the stature of each person in a population is a statistical unit.

The result of an operation carried out on statistical units relating to the individual phenomena which make up a collective phenomenon, is, on the contrary, a *statistical datum*. For example, the number of persons belonging to a population is a statistical datum. The statistical datum therefore measures a collective phenomenon which is the result of the survey of the individual phenomena to which the statistical units refer.

I. 2.6. Magnitudes, Attributes, Modalities

A phenomenon, individual or collective (for instance, a person, an earthquake, a population), is endowed with several characteristics (for instance, the stature or fair color of a person; the duration and violence of an earthquake; the race or the number of inhabitants of a population).

Each one of these characteristics usually offers our senses quantitative degrees (as in the case of the number of inhabitants of a population, or of the height of the body of a person, or of the duration and violence of an earthquake) or qualities (as in the case of color of hair, or race of a population) which may respectively be called *magnitudes* or *attributes* of the characteris-

tic, and will be grouped under the common denomination of *modalities*.

I. 2.7. *Series and Seriations*

If we classify a statistical phenomenon according to the various "modalities" of one of its characteristics or according to those of another phenomenon, and for every class thus obtained we observe the intensity or frequency of the first phenomenon, we obtain a succession of statistical data equal in number to the modalities of the characteristic or of the second phenomenon. According to whether the modalities of this characteristics or phenomenon are qualitative or quantitative, we shall speak of a *series* or, on the contrary, of a *seriation*.

The following are examples of a series (Table A) and of two seriations (Table B) (1).

TABLE A

Geographical Region	Number of Rural Families
Northern Italy. . . .	1.540.081
Central Italy	605.369
Southern Italy. . . .	1.075.134
Islands	518.812
TOTAL	3.739.396

TABLE B

Size of farms in Hectares	Number of Farms in Italy (1930)	Total area of farms (in Hectares)
0.01- 0.50	909.782	199.100
0.51- 1.	581.299	446.359
1.01- 3.	1.273.590	2.398.333
3.01- 5.	532.827	2.092.259
5.01- 10.	492.209	3.482.148
10.01- 20.	253.959	3.535.864
20.01- 50.	106.961	3.188.627
50.01-100.	25.575	1.782.090
100.01-500.	17.559	3.506.073
500.01- and over. . .	3.505	5.620.891
TOTAL	4.196.266	26.251.744

(1) The data in these tables, as well as in several others following, have been taken from S. VIANELLI, *Metodologia statistica delle scienze agrarie*, Vol. I, Edizioni agricole, Bologna, 1954.

Every statistical datum constitutes a *term* of the series or seriation.

In conclusion, we can say that a *statistical series* is a succession of statistical data which measure the intensity or frequency of collective phenomena, distinguished according to the

qualitative modalities of one of their characteristics or of another phenomenon. *Seriation*, on the contrary, is a succession of statistical data which measure the intensity or frequency of collective phenomena, distinguished according to *quantitative modalities* of one of their characteristics or of another phenomenon.

In some cases, the distinction between series and seriation may seem uncertain. The number of marriages according to the days of the week certainly constitutes a series. Let us, however, classify the population or the births of a country according to the years of birth: the succession of data that we obtain may be regarded as a series, notwithstanding that the modalities of the phenomenon, i.e. the years of birth, are expressed numerically. This is because, in reality, the figures which designate the year of birth, do not serve to indicate measurement of the phenomenon but constitute, like the names of the days of the week, the name of the modality. In some countries, in fact, the years are not designated with numbers, but with qualitative epithets (for example, year of the dragon, of the giraffe, etc.).

In order to be able to effect certain calculations, it is sometimes convenient to reduce series to seriations. This can, however, only be done when the qualitative modalities of the phenomenon can be reduced to quantitative modalities. For example, a series representing the number of civil servants or pensioners classified according to their hierarchical rank (qualitative modality) can be transformed into a seriation, since we can replace the hierarchical rank by the corresponding salary, or the number of subordinates that it entails, or the number of years that are necessary, on the average, to reach it, all these constituting quantitative modalities.

Frequency series or *seriations* are those series or seriations which express the frequency assumed by a phenomenon according to the various modalities (qualitatives or quantitatives respectively) of the phenomenon or of one of its characters. For example, the numbers of employees of a firm, classified according to the months of the year, constitute a frequency series. The numbers of firms according to their size, on the contrary, constitute a frequency seriation.

I. 2.8. Different kinds of series

Historical or *chronological* or *time series* are those series whose qualitative modalities refer to divisions of time. *Geographical* or *place series* are those series whose modalities refer to the place. *Qualitative series* are those whose modalities refer to qualities other than time or place.

The following are examples of a time series (Table C), a place series (Table D), and a qualitative series (Table E) :

TABLE C.

Production of grapes in Italy from 1947 to 1954 (in millions of quintals)

Year	Production	Year	Production
1947	53	1951	77
1948	58	1952	74
1949	59	1953	80
1950	61	1954	83

TABLE D.

Agricultural population of the various continents in 1952 (in millions)

Continents	Agricultural population
Europe (excluding the U.S.R.R.)	128
North and Central America	65
South America.	68
Asia	913
Africa.	153
Oceania	4

TABLE E.

*Prices of cheese on the principal Italian markets in 1950,
classified according to the various qualities*

(in It. lire)

Variety of cheese	Price per kilo
Matured fatty	303.90
Dairy Fontina	507.50
Fresh Gorgonzola	260.79
Matured Parmesan	653.89
Roman Pecorine	690.00
Provolone	411.96

Perhaps the most important, anyhow the oldest, among the various kinds of series, are the "historical" or "time" series. They are divided into *static* and *dynamic*.

Those series are called static whose terms remain roughly constant, while those whose terms vary more or less in one direction or in another, are called dynamic. They are divided into *evolutionary* and *non-evolutionary*. The evolutionary series are those in which the terms tend to be modified in a given direction. According to whether this tendency of the phenomenon is towards increase or towards decrease, evolutionary series are respectively known as *progressive* or *regressive*. Non-evolutionary dynamic series are distinguished as *oscillating* or *undulating* according to the smaller or greater degree with which the terms undergo modifications. Among oscillating and undulating series those which manifest a certain regularity in the oscillations or undulations are called *cyclical*, by some statisticians; if the interval between the maxima and minima, and likewise the amplitude of the latter are practically constant they are called *periodical*. Those, on the contrary, which have irregular oscillations and undulations are called *indeterminate oscillating* or *undulating series*, or *erratic series*.

It is not possible to make a sharp distinction between periodical and cyclical series, nor between these and undulating or oscillating series, since their difference lies in the greater or lesser regularity of the variations, so that it is difficult to establish with precision the exact point at which regular variations become irregular, or viceversa.

The oscillations or cyclical variations can naturally occur along with evolutionary tendencies.

Another important distinction is that between *ordered series* and *unordered series*. Ordered series are those whose modalities manifest a natural order of succession. These series are then divided into *rectilinear* and *cyclical series*.

Ordered series which have two extreme modalities, such as, for example, the ranks of the military hierarchy are called rectilinear.

Ordered series for which one cannot speak of extreme modalities, except on the basis of a convention, are called cyclical. For example, to consider January as the first month is the result of a convention; in fact, for the purpose of financial and customs statistics many countries take July as the first month. Once the first term of the series has been fixed, however, the order of the rest of the terms is thereby fixed and cannot be altered.

TABLE F.

Personnel of a group of Italian Government Offices classified according to rank as at July 1st 1949

Ranks	Numbers	Ranks	Numbers
I.	1	VII.	11.230
II.	15	VIII.	16.230
III.	305	IX.	18.739
IV	1.428	X.	9.986
V.	2.764	XI.	5.164
VI.	5.885		

Unordered series are those, the terms of which cannot be regarded as following a natural order of succession, either because they can equally well be put into any order, or because this can be done for some of them only. Unordered series in which the modalities can be put in any order (apart from any

TABLE G.

Temperature in the Rome observatory in 1950
Monthly averages in centigrade

Months	Temperature
January	9.1
February	8.1
March	8.5
April	15.2
May	17.9
June	21.7
July	25.2
August	24.3
September	23.1
October	17.9
November	11.9
December	9.1

TABLE H.

Distribution of the area to be sowed according to the Italian
land survey register of 1929

Crops	Area (in hectares)
Cereals	6.667.422
Industrial crops	268.354
Other crops	1.296.451
Fodder	2.302.366
Fallow with or without pasture	1.200.627
Permanent crops	89.540

convention) and do not present graduations among themselves, graduations, that is to say, serving as transition from one modality to another, are called *unconnected*. Such may be considered a series of persons arranged according to the letter of the alphabet with which their names or Christian names begin. Many series, though not strictly unconnected series, approach such a scheme and can best be treated as such, rather than as any other type; of this type, for example, are series of inhabitants, classified according to religion or nationality; series of agricultural land, classified according to the type of cultivation.

Here are examples of a rectilinear series (Table F), a cyclical series (Table G) and an unconnected series (Table H).

I. 2.9. *Different kinds of seriations*

Seriations are called *continuous* when the interval between two consecutive modalities can be made as small as required, and at the same time, the difference between the two corresponding intensities of the phenomenon can also be made as small as required.

There is discontinuity when the interval between two consecutive modalities cannot be made as small as required (as, for example, in a seriation of agricultural enterprises, classified according to size, when a class of enterprises is missing either on account of a gap in the statistics or because the class does not exist); there is also discontinuity even though the modalities may be continuous, when there is a leap in the intensity of the phenomenon, as can happen in time seriations of imports, due to sudden modifications of customs tariffs. When the discontinuity is found to exist between all the consecutive modalities of the seriation, the latter is called *discrete*; such is the case of the seriation of the number of families, classified according to the number of component members.

Finally, from another point of view, they may be divided into *symmetrical*, *asymmetrical*, *rectilinear*, *parabolic*, *hyperbolic*, *binomial*, etc., according to whether, when expressed graphically, the seriations give rise to figures which approximate to a given type, viz, to symmetrical or asymmetrical figures, or whether they approximate to a straight line, a parabola, a hy-

perbola, a binomial curve etc. We shall see this better later on, when we speak of graphical representation.

I. 2.10. *Variability and Mutability.*

We say that a phenomenon *varies* when it assumes different quantitative modalities (for example, stature, age etc. vary from person to person); we say that a phenomenon *mutates* when it assumes different qualitative modalities (for example, sex, the civil status etc. mutate from person to person). We shall therefore call the capacity of a character to assume different quantitative modalities, *variability*; and we shall call the capacity of a character to assume different qualitative modalities, *mutability*.

We have already distinguished series from seriations; now we shall consider each of them in relation to mutability and variability.

Let us take a series; in this series there is a phenomenon (y) which varies in intensity when the qualitative modalities of another phenomenon (x) mutate. A series thus expresses the values of a phenomenon in function of a *mutable*; by mutable we mean the phenomenon x according to whose qualitative modalities the intensity of y varies: y is called *function* of the mutable x .

Let us take a seriation; in this there is a phenomenon (y) which varies in intensity when the quantitative modalities of another phenomenon (x) vary. A seriation thus expresses the values of a phenomenon in function of a *variable*; by variable we mean the phenomenon x according to whose quantitative modalities y assumes different intensities: y is called *function* of the variable x .

The distinction between magnitudes and attributes, series and seriations, variability and mutability is important in statistics because their quantitative treatment requires different methods.

I. 2.11. *Enumerable and measurable magnitudes, Intensity and Frequency, Amount.*

Intensive magnitudes are usually said to be those which refer to countable phenomena; for example, the number of births or their frequency (natality) in the total of a population.

On the contrary, *extensive magnitudes* are said to be those which refer to measurable phenomena (for example, income, stature, etc.).

Intensive magnitudes are also called *homograde*, since all the statistical units which possess the character, possess it to the same degree, equal to unity; extensive magnitudes, on the other hand, are called *heterograde*, since the various statistical units which possess the character, can possess it to a different degree.

The word *intensity* is however usually taken in a sense which is not in keeping with the above meaning of intensive magnitudes. As a matter of fact "intensity" is commonly called the degree of a measurable phenomenon (for instance, of an earthquake or of the temperature) while the degree of an enumerable phenomenon is called frequency (for instance, the frequency of crimes or births in a population) so that the intensive magnitudes would be those for which one does not speak of intensity but of frequency.

In order to avoid confusion, we think it advisable to speak of *enumerable* and *measurable magnitudes*, and give the word intensity the comprehensive sense of degree of any magnitude, enumerable as well as measurable, including therefore the concept of frequency.

The concept of *frequency* in its turn is sometimes used in the absolute sense of the number of cases in which a certain event occurs; at other times, in the relative sense of the ratio of that number to the maximum number of cases in which the event could happen. When used without qualification, we understand the word frequency in its relative sense.

If a magnitude results from several components, its degree is called *amount* (for instance, the amount of the national income, resulting from the number of income receivers and from the size of the individual income). If the components are all enumerable magnitudes (for instance, the number of earthquake and the number of tremors of the individual earthquake, where the result is the total number of tremors) we shall speak of an *enumerable amount*, while when at least one of the components is a measurable magnitude (as in the case of the national income) we may speak of a *measurable amount*.

CHAPTER II

STATISTICAL SURVEYS

II. 1. PURPOSE OF STATISTICAL SURVEYS

The purpose of statistical surveys is to determine the individual phenomena which constitute the collective phenomenon and, when the individual phenomena are measurable, it also has the purpose of evaluating their amount.

The data usually concern not only the most comprehensive collective phenomenon but also the partial collective phenomena that compose it, e.g. the population census has the aim of determining not only the total number of inhabitants but also their number by age, sex, marital status, occupation etc.

In the survey we may distinguish five phases: 1) general plan of statistical surveys; 2) special plan for a given survey; 3) the collection of data; 4) the counting of the data; 5) tabulation.

II. 2. GENERAL PLAN OF STATISTICAL SURVEYS

In every State, the general plan of statistical surveys is closely connected with the general organization of Statistics and its preparation is the essential task of the head of the statistical bureau and of the consultative statistical body.

The surveys which are planned in a country depend upon several circumstances:

a) The kind of data demanded by the public and by public authorities as well as by international organizations. The former demands have changed with the increasing extension of public and especially State activities. In the past they were concerned essentially with population statistics for administrative and military ends, with judicial statistics for regulating the

administration of justice, with foreign trade statistics necessary for tariff policy, with financial statistics for the purpose of national accounts and level of taxation. For some time, the progress of Statistics was determined to a large extent by the demands of scholars, scientists and research workers, but in recent times it has been caused essentially by the needs of the State and, while in the past such statistics were limited to the collection of data, they are now more and more concerned with statistical research such as the determination of national wealth, distribution of incomes etc. an extension of doubtful opportunity for, if it is true that the State has means superior to those of private scholars and scientific organizations, it does not always have the same objective interest in the search for truth and is more exposed to political pressure. After the first world war, and even more after the second, the planning of State statistical surveys has been strongly influenced by the pressure of international bodies (League of Nations, International Labour Office, United Nations) which are worldwide in scope and have encouraged backward States to organize surveys for which they have no adequate preparation and for which they were obliged to follow more simple, economical, approximate (and sometimes inexact) methods, which for financial reasons tend to become general, depreciating the level of the statistics also in the countries of more ancient civilization.

b) Another circumstance which is important in the general plan of surveys is the traditional activity of the State. In the countries of more ancient civilization, as in Europe, surveys were regularly and rightly organized, and, even if somewhat limited, they were of great importance and made difficult an expansion of statistical returns in other fields; while the new nations, like those of the American continent, concentrated their statistical expansion in the direction of conforming to the modern needs of the public and the State and developed especially Economic Statistics.

c) A remarkable influence on the plan of statistical surveys is exerted by the presence or absence of public organizations which, without being directly statistical in character, will nevertheless assist in contributing to the statistical surveys, such as the system of population registers in many European countries,

the social security bodies, the charitable endowments of Christian and Mohammedan Societies, Labour organization and the co-operative movement.

d) An important negative influence on statistical surveys or at least on the publication of their results may derive from political mistrust, as is the case in the U.S.S.R.

A much debated subject on the organization of statistical surveys is whether they should be centralized or decentralized. Originally, every public department made the surveys that it considered useful for the services in which it was competent. In these circumstances, the need for a co-ordinating body was felt and in nearly every country such a body, usually with consultative powers, was constituted. With the course of time, the tendency prevailed to concentrate the statistical returns in a central office, which in an increasing number of countries is placed, in order to guarantee its independence, under the care of the Chief Minister. In point of fact, the decentralization of statistical surveys among several departments gave rise to conflicts of competence, discrepancies and overlapping between the different statistics, and, what is no less important, gave no possibility to the statisticians of the small offices to reach the higher ranks of the service and obliged the best elements to leave for other activities. On the other hand, centralization has the inconvenience of making the statistical surveys less sensitive to the needs of the administration most interested in them, more difficultly supervised and more likely to be affected by political influences and pressure. A compromise between the two tendencies, which has been proposed in Italy, would be to have, in the different departments, officers detached from the Central Statistical Office. These officers, with the staff of the offices of the departments, would collect the data on behalf of the Central Statistical Office, the different departments also being obliged to submit for approval by the Central Statistical Office the plans of their surveys, as well as of any modification of them.

No less important is the question whether for statistical surveys, the collection and some elaboration of the data is to be made directly by the staff of the Central Office or through local bureaus and employees. In Italy, for example, the Central Statistical Office has the direct responsibility of collection and

itself elaborates the statistical data, while in Spain the greater part of the collection and elaboration is made by provincial officers.

An important question to which insufficient attention is given is that of the opportunity that the Central Offices collect the data through their own staff and not through those of other administrations. It is probable that in the future, the present system, adopted in most and perhaps in all countries, of collecting statistical data through the employees of another department or of local bodies will be judged as ridiculous as we would judge a Minister of War who could pretend to make war with schoolmasters or public employees. At the beginning of the Soviet regime, an autonomous body of statisticians depending on the Central Office was adopted, or at least planned, but it is not known whether this organization was maintained when the Central Office lost its independence and was absorbed by the Planning Department. In Rumania under the presidency of Manuila, the Central Statistical Bureau had decided to have in every municipality at least two employees depending directly on the Central Office, with the job of collecting data. It is not known if this plan, which was still in course of realization, has been completed or on the contrary, discontinued after the occupation of the country by the Russians.

II. 3. PLAN OF INDIVIDUAL SURVEYS

In agreement with the general plan of the statistical surveys, each individual survey requires a special plan, more or less important according to the nature of the survey. Censuses especially need long preparation and careful study, as it is difficult to devise a questionnaire suitable for all cases, so numerous and diversified as individuals, households, farms, factories etc. of a country.

The plan must take account specially of the object of the survey and the means available for carrying it out.

II. 3.1. *Determination of object of surveys.*

A clear definition of the phenomena to be observed is necessary in determining the object of the survey. Many current

concepts are often ambiguous, such, for example, as those of religion, of still birth and of certain occupations. This difficulty is particularly important in agricultural censuses. Definitions of agricultural holdings, land utilization and agricultural population need close examination and a precise definition. On the other hand, it is necessary that the definition be in keeping with the object of the research. For some researches, only live births, for example, must be taken into account, and for others, all births. Sometimes prices to the producer, sometimes market prices, must be considered. It is always advisable not to extend the collection of data beyond what is necessary for the purpose in view: not only would this increase the cost of the survey but also discourage the respondents. Also, it is useless and unsatisfactory to ask for measurements to be given to an unnecessary degree of approximation, for example, asking for the production of oil or wine in litres or that of corn or coal in kilograms.

II. 3.2. Limitations of the survey.

The means provided for the survey are almost always limited as far as time and money are concerned, and some limits of the survey are consequently necessary. Such limitations may concern the scope and the field of observations.

II. 3.2.1. Limitations concerning the scope.

In the first case, the limits may be of precision and specialization. Limits of precision are inherent in agricultural statistical surveys. As it is impossible to collect exact data on the crops, it is usual to base the returns on the personal impression of reporters in villages, districts or provinces, for which it is presumed that they have a special competence. Moreover, the instruments used for some measurements may be chosen to give approximate results, while others give more exact measurements but require more time or greater cost. The limits of specialization consist in choosing among the numerous aspects of the collective phenomenon to be studied, those which have the greatest influence on the results in which we are interested. It is difficult to give general rules for such choice. It depends

essentially on the statistical flair of the observer; it is an art rather than a science. Sometimes, however, past researches will furnish a useful guide to future procedures.

II. 3.2.2. Limitations concerning the field of observation.

These limitations may be distinguished as limits of space, limits of time and limits of the number of observations. For each of these three classes, the limitations can be made by one of three methods: by choosing typical cases, extreme cases (or more generally, compensating ones), or random cases. Often, however, our choice is not free, but is imposed by circumstances. It is often necessary, for example, to collect price-data from cities or important centres and to neglect villages and rural districts. Random choice for space, time and number of observations leads to sampling, a method which has special importance in agricultural statistics. A special course of lectures is devoted to this topic, so that it is not necessary to treat the matter here. Later on, we shall speak of the applications of sampling to other fields (products of factories, public opinion polls) and on that occasion some general notions on sampling will be outlined. It may be added here that this procedure requires a great number of cases among which to choose. When this is not possible, the methods of typical cases or extreme cases are to be preferred.

For analyzing the production, costs and selling prices of various crops, many authors have taken typical farms for study. The choice of date for an agricultural census is that considered typical for the composition of the agricultural population, avoiding periods of heavy emigration and immigration. For the determination of anthropometric characters, as well as the measurement of cattle and other animals, it is advisable, instead of examining a mass of "individuals", to consider those that are considered typical of the race or species and, for human beings at least, careful verification has shown that the results obtained by this method are quite reliable. The method of measuring extreme cases between which an average is taken, is adopted in Italy for the "cadastre" (land register) and everywhere in the measurements of temperature which are taken at the hours of the day which usually show the maximum and the minimum values. Also in this field, records are made of the day of the

year when temperature was at the maximum and of the day when it was at the minimum. The method of extreme cases may be considered as a special case of the method of compensating cases, which is applied, when, having observed one or some cases which deviate from the normal ones in one direction, others are chosen which deviate in the contrary direction, so that a compensation is obtained. Sometimes one extreme only is taken ; for example, for examining the influence of alcoholism and other vices, or of an illness such as malaria, the records are made on the spot where the influence is most serious, in order to be sure that the perturbations do not submerge it.

Limitations of time are justified, and in fact applied, much more frequently for static phenomena (e.g. number of inhabitants) than for dynamic phenomena (e.g. number of births and deaths). In well-ordered statistical returns the collection of data on dynamic phenomena must be continuous, while for static phenomena a periodical collection of data suffices. The length of period is to be determined in agreement with the intensity of the variations, a shorter period being necessary when the variations are most violent.

The three kinds of limitation (by space, time and number of observations) are not independent and are often introduced simultaneously ; for example, in the so-called representative surveys introduced into Statistics by Kiaer, the director of Norwegian Statistical Office and the pioneer of purposive sampling, a fact which has been unjustly forgotten.

The systematic study of the limitations of surveys, to which the Italian treatises have always paid due attention, while it was generally neglected by those of the Anglo-Saxons, has in recent times shown a great development in connection with the importance given to sampling, and especially to purposive sampling, in the public opinion polls and social surveys which gained great favour in America and in England. Consideration has been given principally to the limits of the number of observations, but incidentally also to those of space and time. Many general conclusions thus reached in this field were already known to the statisticians of continental Europe and received confirmation and precision, but many others are new and interesting.

II. 4. THE MODALITIES OF THE SURVEY.

These are concerned with how, when, by whom and with what means the survey must be made.

II. 4.1. *How the survey is to be made.*

It is necessary that the objects of the survey be clear and the instructions available to all those concerned, in the first instance to the enumerators or investigators, and if the survey is directed to persons, to all the population concerned. The questions must be precise, unequivocal, concise, adapted to the minds of the persons concerned and, above all, not suggestive. Experience has shown that the order in which the questions are presented is important; the more ticklish questions should be left to the end. The experience of public opinion polls has been very instructive in this connection and we shall return to this topic when speaking of them in the chapter on sampling (chap. XVI, 3). When there is a tendency to give a round figure, it is advisable to allot it to the middle of the class in which the data are grouped. Careful attention should be paid to the formation of classes or groups of attributes, such as occupation, religion, so that the difference between the classes is always greater than the difference within the classes.

II. 4.2. *When the survey is to be made.*

The date of a survey is very important for some surveys, for example, the census, where it is not easy to find a solution suitable for all parts of the country and for the different categories of the population. In making researches into primitive populations it is important to contact them during a favorable period, as they tend to connect the happenings with the arrival of a stranger; the traditional customs and taboos must not be disturbed.

II. 4.3. *By whom the survey is to be made.*

To ensure the success of a survey, good direction is necessary as well as good and efficient staff, and, if the survey is

directed to human beings, the willing co-operation of the respondents of the survey. It is not easy to realize the last two conditions in under-developed countries and it is a general impression that statistical data beyond the capacity of the people concerned are now collected for international bodies. Many surveys, such as those on health, sexual matters, morality, illegal practice and evasion of laws present special difficulties: they must be carried on only if the strictest secrecy is guaranteed, and they presuppose special tact and competence in the investigators.

II. 4.4. The means for carrying on the survey.

These means differ according to whether the collection of the data is continuous, as is usually the case for dynamic phenomena, or discontinuous, as is the case for static phenomena. In the first case, registers are usually set up; in the second, questionnaires or cards are used. In the course of time, the means or instruments used in collecting data have become more numerous, varied and complex, especially in some kinds of research (such as medical, physical or psychic measurements).

Precision, reliability, coverage, timeliness and comparability of data are always to be aimed at in drawing up the plan of the survey. Some of them, particularly the last two, depend in great part on the kind of collection and tabulation of the data, but they also depend partly on the determination of the object, the limitations and the modalities of the survey, of which we have just spoken.

II. 5. COLLECTION OF DATA.

Some distinctions between kinds of statistical surveys are opportune, with special reference to collection of data.

Preliminary and definitive surveys: Complicated statistical surveys such as the census require a preparatory survey in order to determine such questions as the number of questionnaires to be distributed, the number of enumerators required, etc., but the majority of inquiries do not need such precautionary measures. Independently of their administrative complications, inquiries which present some points for which difficulties are fear-

ed without it being possible to define them exactly, may be preceded by "pilot surveys", which consist in an attempt to carry out the final survey on a reduced scale.

Occasional, periodic or continuous collection of data : The census is a typical periodic survey ; collections of vital statistics (births, deaths etc.) are typical continuous surveys in well-organized countries. Special inquiries, such as, in Italy, the agrarian survey in 1885 and other inquiries on the conditions of peasants in Southern Italy and Sicily in 1906-07 and a recent one on unemployment, are determined by special circumstances and are made when these occur.

Automatic and reflex collection of data : Automatic collection arises when the information is brought to the knowledge of the authorities by the persons concerned, or their representatives or relatives, as, for example, for births, marriages and deaths ; on the contrary, reflex collection of data arises when the information is collected by persons specially charged by public or private inquirers with this duty. Usually the automatic collection is continuous and the reflex is periodic or occasional ; sometimes, however, reflex collection of data may be continuous ; for example, the collection of data relating to persons crossing the frontier of a State.

Extensive or intensive surveys : In contrast with the usual statistical surveys, which are described as extensive, there are surveys called intensive, where the object is the thorough study of a special or limited group of persons or objects, and where advantage is taken not only of quantitative but also of qualitative information. A typical example of an intensive survey is given by the family monographs of Le Play.

Intensive surveys in the past were confined to a small number of groups, often to a single group, while now they are frequently extended to a large number of groups and, from this point of view, they acquire a more pronounced statistical character.

Direct and indirect collection of data : Sometimes, not having the possibility of examining or measuring the intensity of the phenomenon which is the object of interest, we collect data

on another phenomenon which justifies us in drawing inferences concerning the original phenomenon. In the past, for example, instead of enumerating the population, the number of chimneys were counted and the population was estimated on the basis of the average number of persons per family, assuming that the number of chimneys corresponded to the number of families. For primitive peoples, a similar method - based on the number of huts - has been adopted. Indirect surveys, properly so called, must be distinguished from conjectural statistics in which no new survey is made, but the data already collected for a certain object are used for evaluating another phenomenon as, for example, when the statistics of tax assessments are used to estimate the national income.

Complete, incomplete and over-complete: surveys: Complete surveys are those in which each individual phenomenon is observed once only: incomplete surveys are those in which some of the individual phenomena are not observed. Surveys may be called over-complete when all the individual phenomena are observed and at least some of them more than once. This happens, for example, when the number of children is asked for, both on the death of the father and on the death of the mother, or when, on the occasion of each new birth, the number of male and female children of the sibling is asked for and used for the computation of the sex-ratio.

Representative and non-representative surveys: This distinction is possible only for over-complete or incomplete surveys. Representative surveys are the incomplete or over-complete surveys in which the different modalities are observed with the same frequency as in complete surveys; otherwise the survey is said to be unrepresentative. The lack of representativeness may depend on systematic causes and in that case the survey is said to be biased; or it may depend on accidental causes, i.e. on accidental errors of observation or on random errors dependent on the small number of observations. It is important to remember that surveys may be representative for one character but not for others. This fact greatly reduces the utility of purposive sampling.

II. 6. THE COUNTING OF THE DATA

In this phase of the survey we may distinguish three operations, the enumeration, the classification and the material operation of counting.

II. 6. 1. *Enumeration.*

The object of this is to evaluate the number of individual phenomena observed. This is necessary, for example, when the declarations are not compulsory and useful even when they are compulsory for giving a general idea of the size of the inquiry and as a means of checking the number of cases included in the different categories established in the following operations.

II. 6. 2. *Classification.*

This consists in the grouping of the data according to the modalities of various characteristics of the phenomenon. Thus, inhabitants of a country are classified by the modalities of sex, age, civil status, occupation, residence, place of birth etc. Classification may be made on the basis of a single character or by combination of two or more characters. It is always possible to pass from a classification by combined characters to a classification by a single character, but not vice versa. Classification by a combination of characters is therefore to be preferred, but it is necessary not to make too complicated combinations, otherwise the number of cases in each combination may become too small and the general aspect of the statistical phenomenon cannot be grasped. These observations concern the definitive classification, but it is wise to begin with a detailed classification, afterwards grouping the classes as may be convenient. This is especially the case for classification of a territorial nature, in order to give the possibility of forming natural zones or regions, drawn up according to the character of the phenomenon studied, rather than the artificial zones created for administrative purposes.

It is obvious that the criterion for distinguishing the classes may be different according to the object and nature of the investigation. Account must be taken of the details necessary and of the limits which may be of interest, for example, the limit for

income tax exemption, the limits fixed for acceptance for military service etc. It is very useful for the statistical elaboration and visual presentation of the data, that the different classes have the same interval or amplitude, but this is not possible when the frequency is very different for the various modalities, as, for example, in the distribution of incomes and of wealth.

II. 6. 3. *Material operations of counting.*

General statistical surveys are now usually treated by mechanical means, where punching machines prepare punched cards which are then passed, after verification, to the sorting and counting machines and finally to the tabulating machine. Mention may be made of the less progressive procedure, still used in the smaller surveys of the entries by "ticks" on prepared paper; this procedure may be recommended only when the number involved is not too great. Counting machinery has also been used, especially by the French official statistics, for extracting the data directly from the questionnaires. Usually, however, both operating by hand by machine, the transcription of data on to cards, sometimes of different color and size, precedes the material operation of counting.

II. 7. *Tabulation.*

Tabulation is an important phase of statistical surveys for bringing to light all the aspects of the data. Tabulation is a simple operation to an experienced statistician, but to a layman it often presents difficulties.

The simplest statistical table consists of a column of figures, giving on the left hand side, the list of modalities of the phenomenon to which the figures refer: it is called *simple tabular statement*; e.g. a statement showing the months of the year in the first column and the figures of marriages occurring in each month in the second column. Examples of simple tabular statements can be found at pages 340 (table A), 342 (table D), 343 (table E), 374. The first column on the left is called *them other column* or *stub*, and captions at the head of the columns form *column headings*.

If the mother column contains the combined modalities of two or more characteristics, the *tabular statement* is said to be *complex*.

The following is an example of a complex tabular statement:

Distribution of the wooded area in Italy

Height of the zone	Kind of wood	Hectares
Mountain	Resinous wood	862.194
	Broad-leaved tree woods	650.445
	TOTAL . . .	1.512.639
Hilly zone	Resinous wood	131.030
	Broad-leaved tree woods	354.625
	TOTAL . . .	485.655
Plain	Resinous wood	46.315
	Broad-leaved tree woods	57.093
	TOTAL . . .	103.408
TOTAL . . .	Resinous wood	1.039.539
	Broad-leaved tree woods	1.162.163
	TOTAL . . .	2.101.702

If, instead of having only one column of figures, there are two or more columns giving the figures corresponding to the various modalities of a phenomenon, we have a *statistical table proper*. It may also be *simple* or *complex* according to whether both the *mother column* and the *column headings* contain modalities of only one characteristic or, on the contrary, at least one of them contains modalities of more characteristics.

An examples of a simple statistical table in the table I.

A complex tabular statement can always be transformed into a statistical table. Thus the data of the complex statement on the distribution of wooded areas given above can be presented in a more convenient form in the simple statistical table given at page 387.

Other examples of simple statistical tables can be found on pages 367, 385, 390, 409.

An example of a complex statistical table is the table II. It gives the combination of two characters (time, kind of machinery) in the headings in function of the modalities of a third characteristics (geographical region) listed in the mother column.

Another exemple can be found at page 368.

It is necessary to distinguish the statistical tables proper from the *folded tabular statements* as well as from the *compounded tabular statement*.

If the list of the modalities considered in the mother column is rather long, it may be advisable to fold the tabular statement (as f.i. is done at page 343, table C), but this does not alter the fundamental characteristics of the simple tabular statement, which is that of giving the figures for a single phenomenon in function of the modalities of another phenomenon.

TABLE I

*Farm machinery of U.M.A. classified by geographical region
in 1938 and 1948*

Geographical region	Farm machinery in Italy	
	1938	1948
Northern Italy.	41.366	72.808
Central Italy.	12.451	22.648
Southern Italy.	6.258	13.104
Islands	3.314	7.120
ITALY	63.389	115.680

If, in function of the modalities of the same phenomenon, we write the figures concerning two or more other phenomena (as f.i. is done at table B, page 340), we obtain a compounded tabular statement which differs from a statistical table proper, because it does not give, in function of the modalities listed in the mother column, the figures for the modalities of another

TABLE II

Farm machinery of U.M.A. classified by their geographical region and the kind of machinery in 1938 and 1948.

Geographical Region	1938		1948	
	Tractors	Other machinery	Tractors	Other machinery
Northern Italy	26.591	14.775	41.096	31.712
Central Italy	5.870	6.581	7.766	14.882
Southern Italy	3.122	3.136	3.590	9.514
Islands.	1.381	2.640	1.933	5.187
ITALY . . .	36.964	27.132	54.385	61.295

phenomenon but the figures for two distinct and not additive phenomena (in the example on pag. 340, number and area of the farms).

Also for the statistical table proper we may fold our simple statistical table when the list of the modalities contained in the mother column is long and we may also compound several simple statistical tables when, in function of the modalities of the mother column, we express in the column headings the modalities of two or more not additive phenomena.

Here is an example of a compounded simple statistical table (See table III).

Lastly we may also have folded and compounded complex statistical tables. The following (See table IV) is an example of a compounded complex statistical table.

The folded and compounded statistical tables are useful as they offer a maximum of the combination of the data collected, but at the same time it is difficult to grasp from them the relations between the various phenomena. For that reason they are used more especially in the appendices of the official statistical reports, but more simple tables are advisable and are generally used in the text of scientific research as well as in those referring to practical investigation.

TABLE III

Rural families classified by geographical and occupation of the head of family on 21st-4-1936

Geographical	Rural families		Number of amilies whose head is					Number of members of the families whose head is				
	Number	Members	Non- operating farmer	Operating farmer	Share- worker	Worker	Employee	Non- operating farmer	Operating farmer	Share- worker	Worker	Employee
Northern Italy	1,540,081	8,136,579	52,632	895,475	160,914	424,565	6,495	253,195	4,738,258	1,232,619	1,883,481	29,026
Central Italy	605,369	3,409,938	21,652	220,687	216,811	140,839	5,380	88,270	1,125,537	1,594,681	581,875	19,575
Southern Italy	1,075,134	5,063,953	49,476	562,842	94,724	365,918	2,174	202,958	2,775,622	520,967	1,553,799	10,597
Islands	518,812	2,278,288	36,416	195,311	76,878	208,522	1,685	138,402	879,358	360,438	892,873	7,217
ITALY	3,739,396	18,888,758	160,176	1,874,315	549,327	1,139,844	15,734	682,835	9,518,775	3,708,705	4,912,028	66,415

TABLE IV

Active population employed in agriculture classified by professional situation, on 21st-4-1936

Professional situation	As a whole		In		
	total	of whom family-assistants	farming	live-stock	forestry
A. Farmers :					
a) land-owning	2.872.648	1.392.703	2.796.504	74.468	1.676
of whom operators	2.704.769	1.357.534	2.638.399	65.171	1.119
b) non land-owning	863.106	451.474	862.601	414	91
of whom operators	798.628	428.750	798.198	354	76
b ₁) usufructuaries	26.154	5.000	26.109	24	21
of whom operators	12.406	4.668	12.372	12	21
b ₂) emphyteutes	40.133	20.462	40.133	—	—
of whom operators	37.790	19.912	37.790	—	—
b ₃) tenants	796.819	426.022	796.359	390	70
of whom operators	748.433	404.170	748.035	342	55
c) sundry	709.939	407.695	708.745	1.165	29
of whom operators	689.263	398.956	688.084	1.170	29
TOTAL FARMERS	4.445.693	2.251.882	4.367.850	76.047	1.796
OF WHOM OPERATORS	4.192.660	2.185.240	4.124.681	66.675	1.304
B. Crop-sharers	1.787.601	1.204.235	1.784.940	2.612	49
C. Mixed figures					
a) farmers and day workers	103.211	37.934	103.114	85	12
b) farmers and workers by annual contract	683	305	660	23	—
c) farmers and crop-sharers	11.944	5.024	11.812	132	—
d) day workers and farmers	21.977	—	21.977	—	—
e) workers by annual contract and farmers	302	—	275	27	—
f) workers, crop-sharers and farmers	760	—	760	—	—
TOTAL MIXED FIGURES	137.877	43.263	138.590	267	12
D. Workers :					
a) crop-sharing workers	141.377	—	140.235	1.046	96
b) day workers	1.817.283	—	1.761.866	5.063	50.354
of whom unskilled	1.731.770	—	1.727.729	873	3.168
c) by annual contract	381.785	—	332.520	46.284	2.981
of whom unskilled	164.812	—	162.377	2.306	129
non-agricultural	16.073	—	14.289	448	1.336
of whom unskilled	458	—	258	58	141
TOTAL WORKERS	2.356.518	—	2.248.910	52.841	54.767
of whom unskilled	2.038.417	—	2.030.599	4.284	3.534
E. Estate managers and employees :					
a) agricultural	17.388	—	16.960	142	286
b) non agricultural	4.022	—	3.582	198	242
TOTAL MANAGERS AND EMPLOYEES	21.410	—	20.542	340	528
F. unclassified	5.965	5.955	5.942	18	5
GRAND TOTAL	8.756.064	3.505.335	8.566.782	132.125	57.157

When in a simple statistical table all the modalities of the characters listed in the mother column and the others in the column headings are comparable either because both the characters are quantitative or because, being qualitative, they present the same or analogous modalities, the table is said to be a *double-entry table*. When in a complex statistical table the three or four characteristics present comparable modalities, the *table* is said to be of *triple or quadruple entry*. Tables with a higher number of entries are not desirable.

Tab. V is an example of double-entry table; another example can be found at page 393.

By combining the measurements of the cephalic index of the fathers and children given in the above double-entry table with respective measurements concerning the mothers, the triple-entry table given of page 368 is obtained. (Table VI).

TABLE V

Children of immigrants in the United States classified by their cephalic index and that of their fathers

Cephalic index or the children	Cephalic index of the fathers			
	up to 79,9	80,0 to 84,9	85,0 and more	total
up to 79,9	57	134	39	230
80,0 to 84,9	118	613	247	978
85,0 and more	54	365	208	627
TOTAL . . .	229	1.112	494	1.835

It is generally advisable to have recourse to some devices, to make the tables more easily understood. When the columns are numerous, they should be numbered consecutively and the same applies to the lines. It is also useful to give at the end of the lines and at foot of the columns, the totals of the lines and columns respectively; these are called *marginal distributions*

TABLE VI

Children of immigrants in the United States classified by their cephalic index and those of their fathers and mothers

Cephalic index of the children	C e p h a l i c i n d e x o f t h e f a t h e r s											
	79.9 Ceph. i. of the mothers		80.0 84.9 Ceph. i. of the mothers		85.0 Ceph. i. of the mothers		Total Ceph. i. of the mothers					
	— 79.9	80.0 84.9	— 79.9	80.0 84.9	— 79.9	80.0 84.9	— 79.9	80.0 84.9				
up to 79.9	27	23	7	82	34	18	12	21	6	73	126	31
80.0 to 84.9	35	64	19	372	144	97	38	165	44	217	601	160
85.0 and more	7	31	16	210	43	112	28	121	59	78	362	187
totals	69	118	42	664	221	227	78	307	109	368	1089	378
TOTAL		229		1,112				494			1,835	

or marginal frequencies. It is advisable to use different fonts (roman type, italics, heavy type, etc.). If, in a complex table, there is a total figure and another figure which forms part of the total, it is better to give the latter figure in brackets or in smaller type. When possible, the headings should be given completely and horizontally; but, if they are too long, they can be abbreviated, written vertically or indicated by a letter (A, B, etc.) which can be explained at the foot or at the head of the table.

CHAPTER III

VISUAL PRESENTATION OF STATISTICAL DATA

The purpose of visual presentation is to illustrate, by the use of figures, solids, or other means, the distributions of statistical phenomena and the variations that these present according to the variations and mutations of other phenomena.

The most important of all the systems of visual presentation is certainly graphical representation.

III. I GRAPHICAL REPRESENTATIONS

III. I.1 *Diagrams.*

The most widely used graphical representations are *diagrams* and *cartograms*.

Diagrams are figures that illustrate, by means of geometrical elements, the quantitative modalities of a phenomenon in relation to the quantitative or qualitative modalities of other phenomena.

The fundamental principles of their construction derive from the norms of analytical geometry, which establish the correspondence between elementary geometrical entities (points, lines, areas, angles) and numerical entities (measurements of the the modalities and of quantitative phenomena).

Diagrams can first of all be divided into *simple* and *multiple* according to whether they serve to represent one or more than one phenomenon as a function of the same complex of other phenomena (quantitative, qualitative, or quantitative and qualitative together).

Simple diagrams can be constructed to represent :

A) A quantitative phenomenon as a function of a qualitative phenomenon ;

A') A quantitative phenomenon as a function of another quantitative phenomenon.

B) A quantitative phenomenon as a function of two qualitative phenomena ;

B') A quantitative phenomenon as a function of one qualitative and one quantitative phenomenon ;

B'') A quantitative phenomenon as a function of two other quantitative phenomena ;

C) A quantitative phenomenon as a function of three or more phenomena.

A) *A quantitative phenomenon as a function of a qualitative phenomenon.*

The quantitative modalities of the function are represented on a plane by means of a corresponding number of geometrical figures, of which a linear dimension or an area expresses the measurement of the respective modality. For such a purpose one can make use of rectilinear segments, of rectangles or triangles having equal bases, of squares, or of circles or sectors of circles etc., in such a way that the lengths, the heights, the sides, the areas, the extensions etc. are respectively proportional to that measurement.

Let us suppose that we have to measure the agricultural population of the various continents in 1952 (See page 24, table D). We can for this purpose, from a single line of origin trace segments of length proportional to the respective quantity of population living in each of the continents (fig. 1). This system is called *linear diagram*.

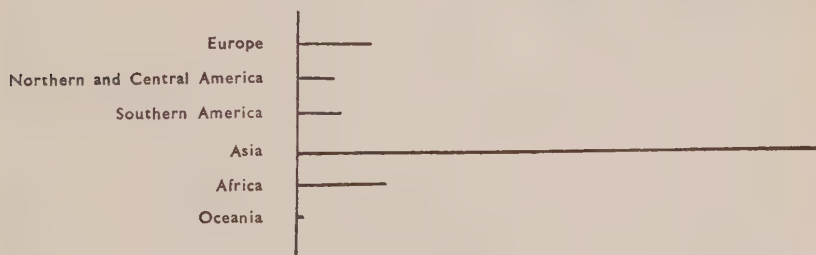


Fig. 1 — Agricultural population in the various continents in 1952

Areograms, or representations by means of areas, are however clearer. Although for their construction recourse may be had to any type of area, the most widely used figure is the rectangle because of the greater facility with which the eye can estimate the differences of intensity of a phenomenon represented by a series of rectangles of equal height and of different length, technically known as *horizontal-bar chart*, or by a series of rectangles of equal base and of different height, technically known as *vertical-bar chart* (fig. 2). The *areograms composed of rectangles* are also suitable for the representation of phenomena whose amount is the product of the intensities of two component phenomena, making one dimension proportional to the intensity of one of

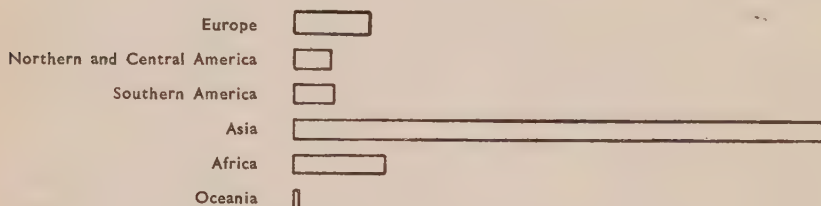


Fig. 2 — Agricultural population in the various continents in 1952

these and the other proportional to the intensity of the other, so that the areas of the resulting rectangles are proportional to the amounts of the resultant phenomenon which is the product of the two component phenomena.

For example, if we wish to represent the annual production of wheat, we can construct as many rectangles as there are years to be indicated, each rectangle having a base proportional to the area of land devoted to cultivation of wheat in the respective year, and a height proportional to the average output per hectare obtained in the same year; the area of each rectangle will thus be proportional to the production for the respective year.

A widely used type of graphical representation is that known as *pictograms* or *ideographs*; these consist of diagrams with shapes similar to one another and recalling to the mind the object to which they refer; for example, the human form to represent population, the shape of barrels to represent wine production,

of ships to represent maritime traffic or the consistency of the merchant navy.

The number of cases or the amount of the quantitative phenomenon is presented as proportional to the length or height, area or volume of the shape, preferably to area as we may consider that the observer is influenced by the area of the shapes rather than by their length, height or volume.

In order to avoid the misunderstandings that can certainly arise, it is preferable to adopt the *system of repeated shapes* in which every figure represents a fixed number of cases or a fixed amount of the phenomenon (for example, a thousand inhabitants, a hundred hectolitres of wine, a thousand tons of gross tonnage, etc) and is repeated as many times as necessitated by the total number of cases or total amount of the phenomenon to be represented, a fraction of the fixed number of cases or amount being represented by a section of the shape. These shapes are equally spaced in a horizontal or vertical line, thus giving a line or column whose respective length or height is proportional to the total number of cases or amount of the phenomenon.

For example, if we wish to represent the wine production in Italy from 1945 to 1951, as given in the following table, we can indicate a quantity of 10,000 hectolitres by a wine-barrel, and if we wish to indicate each year's production we can put one after the other as many equally spaced barrels as there are tens of thousands of hectolitres produced in the respective year, repeating the same operation for each year on consecutive lines proceeding downwards. The thousands of hectolitres that do not reach ten are represented by a section proportional in size to a barrel (fig. 3).

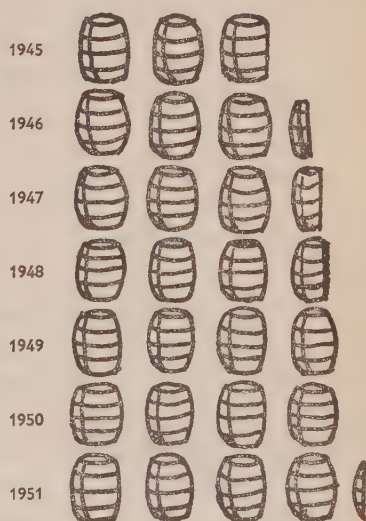


Fig. 3 — Production of wine in Italy

Production of wine in Italy from 1945 to 1951
(*Annuario statistico italiano*, 1952, page 449)

Year	Wine production (in thousands of hectolitres)
1945	29
1946	33
1947	35
1948	39
1949	40
1950	40
1951	41

A') *A quantitative phenomenon as a function of another quantitative phenomenon.*

The method most commonly used to represent a quantitative phenomenon as a function of another quantitative phenomenon is that of *orthogonal Cartesian coordinates* (so called by reference to Descartes, who was not in fact the first to propose them but who was responsible for their adoption in analytical geometry). The logical foundation of such a method is the indication of the various values of a phenomenon by means of different points on a plane, each of the points corresponding to a determined value of the phenomenon.

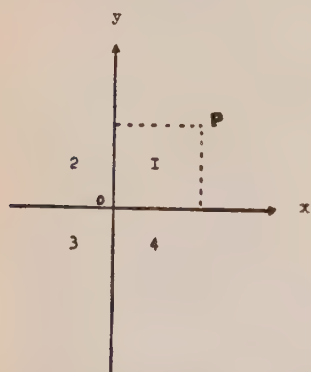


Fig. 4 — The four quadrants in the Cartesian orthogonal system

These points in the Cartesian system are determined by means of reference to two straight lines which cross one another, and which, if they form a right angle, give rise to the *Cartesian orthogonal system*. This is very useful since any point on the plane can be determined by the use of very simple means and it easily understood.

The two perpendicular straight lines divide the plane into four right angles or *quadrants* (fig. 4). Any point on the plane necessarily belongs to one the four quadrants and its position is perfectly determined when the distance is known which separates it from the two straight lines limiting the quadrant in which it

is situated. Generally all the quadrants are not used but only that which is indicated by the number 1 in figure 4; use of the others is made only if one or both of the phenomena present negative values. The two lines of reference are called *co-ordinate axes*; their point of intersection *o* is called the *origin*; the perpendicular from *P* to horizontal axis is called the *ordinate* of *P*, and is naturally parallel to the vertical axis; the perpendicular from *P* to the vertical axis is called the *abscissa* of *P* and is naturally parallel to the horizontal axis.

Every possible point in the quadrant thus has its abscissa and its ordinate. All the possible abscissae are parallel to the horizontal axis, which is therefore called the *axis of the abscissae* or, because of the letter which is commonly used to indicate it, the *x-axis*. All the possible ordinates are parallel to the vertical axis, which is therefore called the *axis of the ordinates* or the *y-axis*. Generally the *variable*

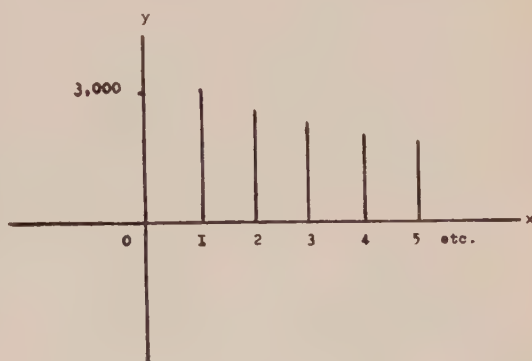


Fig. 5 — Ordinate diagram

is represented on the *x-axis* and the *function* on the *y-axis*. If it is borne in mind that a point is perfectly determined by an *x-axis* and an ordinate, the mechanism of this graphical representation is easily explained.

If, for example, we have to represent the distribution of the agricultural population as a function of their different ages (fig. 5), we shall take on the *x-axis*, starting from 0, consecutive segments proportional in length to the interval of time considered. In our case, since it is a question of years, we shall take segments which are all equal. On the *y-axis*, also starting from 0, we shall take consecutive segments proportional in length to the quantity of agricultural population comprised in the successive classes of age under consideration.

If the agricultural population of one year of age is 3,000, this datum will be represented by the point of intersection of

the ordinate raised from the point 1 on the x -axis and the abscissas perpendicular to the point 3,000 on the y -axis. In practice it is sufficient to raise only the ordinate from the point on the x -axis corresponding to the value of the abscissa, in our case 1, for a length equal to that corresponding to the value of the same ordinate on the y -axis, in our case 3,000.

In order, therefore, to represent the distribution of the agricultural population according to their different ages, it will be

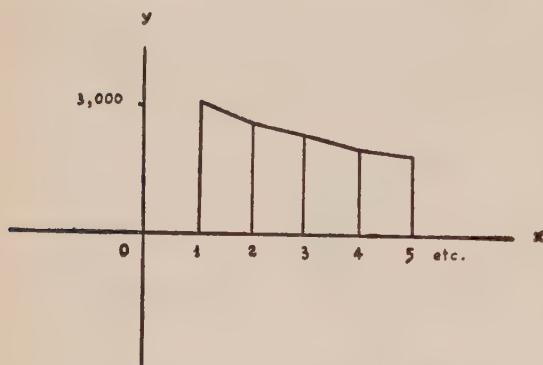


Fig. 5 bis. — Joined ordinate diagram

sufficient to raise, from the extremes of the segments taken on the x -axis, perpendiculars (ordinates) proportional in length to the quantities of agricultural population according to their different ages. In this way we obtain an *ordinate diagram*.

In practice, however, it is usual to mark only the extremes of these ordinates and to join them with a line, thus giving rise to a *joined ordinate diagram*, which the Americans call a *line-diagram chart*, and which shows the development of the phenomenon more clearly (see fig. 5 bis).

The line which results from the joining of the successive ordinates is in reality a broken line, but it is commonly called a *curve*, since, in the very frequent case in which the variable is continuous, the appearance of a broken line is due only to the arbitrary division of the variable into discontinuous classes, while if they were continuous, that is to say if each class were of infinitesimal extent, the extremes of the relative ordinates would be consecutive and would form a curve.

In order that the presumable curve corresponding to a broken line may be deduced with a minimum of arbitrariness, there are special instruments called *curve-rulers*, which consist of flexible steel or plastic blades which can be fixed in the points

given, viz, at the extremes of the ordinates, and which can be curved regularly between the points given. Although the arbitrariness is not entirely eliminated, since these steel blades only trace curves of one given type amongst all the infinite types possible, they serve to eliminate the abrupt discontinuities of the line and to make it independent of the drawer. (fig. 6).

The determination of the scales to be adopted for the measurement of the variable and of the function introduces, however, certain elements of arbitrariness. In any case, once a particular scale has been adopted, the same scale must be maintained unchanged throughout the graph and, when it is a question of data referring

to classes, the lengths of the respective abscissae must be proportional to the size of the classes that they represent.

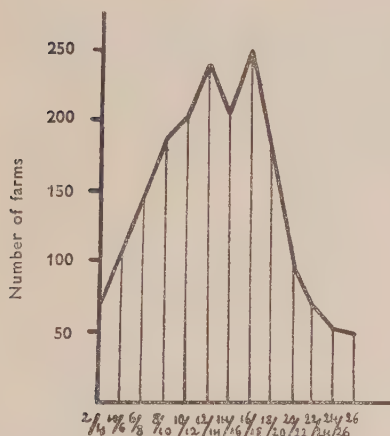


Fig. 7 — Number of farms according to size (Rhine Reclamation Scheme)

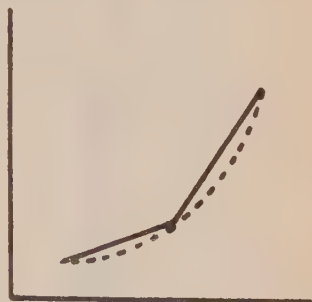


Fig. 61 — Comparison of a joined ordinate diagram with the respective curve traced with the curve ruler

In a seriation, moreover, the size of the classes can be all of equal interval as in the case of the farms comprised in the Rhine Reclamation Scheme classified according to their size in hectares (See Chap. IV 2.6.) or of different intervals as in the case of agricultural enterprises in Italy in 1940 classified according to their size (See Table B, Chap. I 2.7) for the graphic representation we have grouped the first two and the last two classes). If we have to represent the two seriations graphically, in the first case we shall mark off

equal segments on the x -axis (fig. 7), and in the second case unequal segments proportional to the different sizes of the classes (fig. 8).

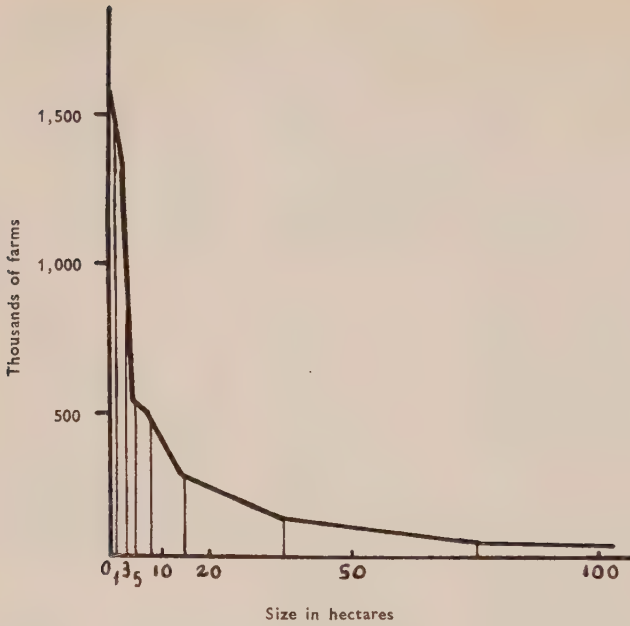


Fig. 8 — Frequency polygon. Number of farms according to size. Italy, 1940

In order that erroneous appreciations of the relative importance of the variations of the function may be avoided, the ordinates indicated in the graph should always start from *nought*, even if the variations only concern very high values of the ordinates.

The different impression of the importance of the variations given by figures 9 and 10 should be observed. Although they

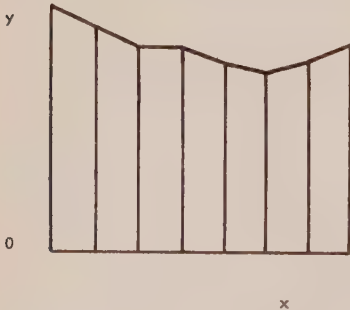


Fig. 9 — Correct representation: ordinates starting from nought



Fig. 10 — Representation of same data as in fig. 8: Ordinates starting from value *n*

represent the same curve, in figure 9 the ordinates begin from nought, while in figure 10 they begin with a higher value n of y .

With orthogonal co-ordinate diagrams the mathematical functions which are most frequently used in Statistics can be adequately represented.

In fact a linear or first degree equation, represented in a Cartesian co-ordinate diagram gives rise to a straight line. It

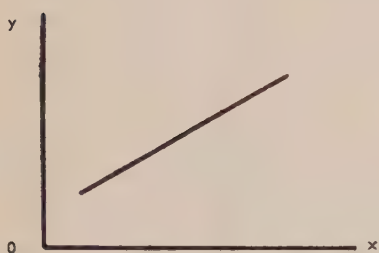


Fig. 11 — Ascending straight line

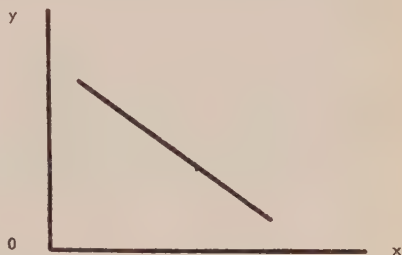


Fig. 12 — Descending straight line

corresponds to the formula $y = a + bx$ where y is the function, x the variable, and a and b the *constants* or *parameters*. The straight line will be ascending (fig. 11) or descending (fig. 12) according to whether b is positive or negative.

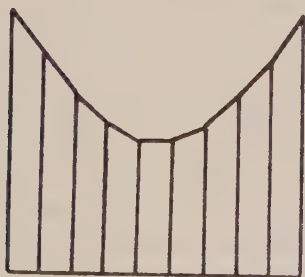


Fig. 13 — Representation of an ordinary parabola



Fig. 13 — Curve of hyperbolic type

A second degree equation, corresponding to the formula $y = a + bx + cx^2$, gives rise to curves called *ordinary parabolas* (fig. 13).

A third degree equation, corresponding to the formula $y = a + b x + c x^2 + d x^3$, gives rise to *parabolas of the third order* and so on.

Another important category of curves is constituted by *logarithmic* or *exponential curves*, which correspond to the formula $y = ba^x$. In terms of logarithms, this is transformed into the linear equation $\log y = \log b + x \log a$.

Important also are *curves of hyperbolic type*, corresponding to the formula $y = \frac{a}{x^m}$, where m is real and less than nought.

In fact incomes are usually distributed in accordance with a curve of this type (fig. 14).

The *curve of accidental errors* or *Gaussian curve*, or *normal curve*, of which we shall have occasion to speak repeatedly, is a symmetrical exponential curve corresponding to the formula $y = K e^{-bx^2}$, where e represents the base of the natural logarithms. Its bell-like appearance is shown in figure 15.

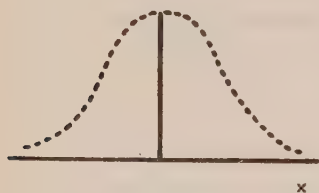


Fig. 15 -- Normal or Gaussian curve

Sinusoidal curves, defined by trigonometric equations with sines and cosines are suitable for the representation of periodic phenomena.

If we represent a group of farms classified according to their size, the value corresponding to each class of size is nothing but the (absolute) frequency possessed by the same class. In this case, the figure formed by the x -axis, the broken line, the first and the last ordinate is called *frequency polygon* (fig. 8). One can however represent each of these frequencies, not by means of an ordinate raised from a more or less arbitrary point of the segment which represents the interval of variation of the variable, but by means of a rectangle which has as its base the said segment, and of a height such that the area of the rectangle is equal to that frequency. The figure will thus appear as a series of adjacent rectangles standing on the same straight line. It is called a *histogram*.

An example of a histogram is given in figure 16 constructed on the basis of the data used for figure 8. When this system is

used, the total frequencies of each class are divided by the size of the same class; this gives the unitary mean frequencies within the ambit of each class and these are then used for the graphical determination of the height of the ordinate of the rectangle. In this case the frequency is proportional to the area of the corresponding rectangle, while in the former case it was proportional to the height of the corresponding single ordinate. If the bases of the rectangle, that is the sizes of the classes, are equal, the

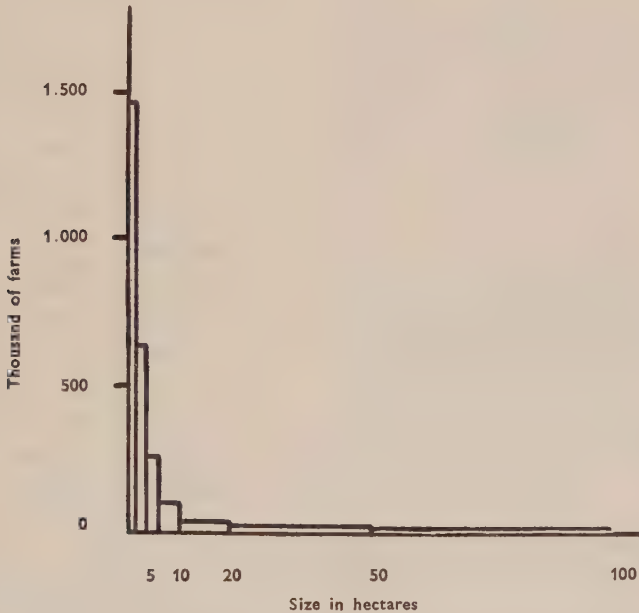


Fig. 16 — Histogram, Number of farms according to size (same data as in fig. 8)

frequencies are then proportional to the heights of the corresponding rectangles; when the bases are unequal however, that is when the classes are of different sizes, then this does not happen. Hence the danger of using ordinate diagrams and the advisability of using histograms in their stead.

In our example, the fig. 16 gives a correct representation and the fig. 8 a wrong representation of the phenomenon.

Other diagrams to which the statistician must frequently have recourse in the case of a quantitative phenomenon as a

function of another quantitative phenomenon are single or double *logarithmic charts*. The former are generally constructed by taking as abscissae the natural values of the variable and as ordinates the decimal logarithms of the function. They are espe-

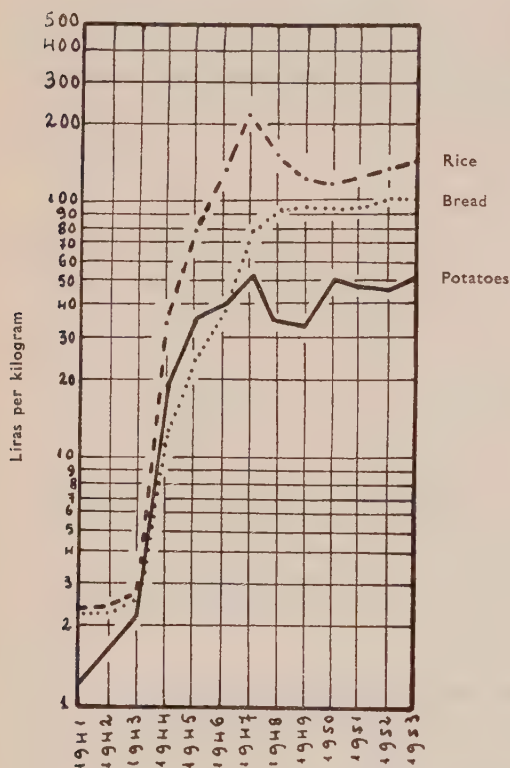


Fig. 17 — Single logarithmic chart. Prices of some foodstuffs in Italy

cially adopted when the function oscillates widely, since the logarithms of the numbers increase less rapidly than the numbers themselves. Single logarithmic charts also have the property of giving an exact image of the relative importance of the variations, which is sometimes more interesting than that of the absolute variations, as indicated by the usual diagrams. As an example let us use this system to represent (fig. 17) the trend of the retail prices of some foodstuffs in Italy from 1941 to 1953. The data of the following table are taken from the *Compendio Statistico Italiano* of 1954, page 351.

Double logarithmic charts, that is those having logarithmic scales both for the abscissae and for the ordinates, are principally useful for phenomena following the formula $y = kx^a$, which for x real and greater than nought give curves of parabolic type when normal scales are used, and for x less than nought curves of hyperbolic type. If we use double logarithmic diagrams, we have $\log y = \log k + a \log x$, so that if we make $X = \log x$, $Y = \log y$, $K = \log k$, we obtain $Y = K + aX$, which means that the parabolic curves and hyperbolic curves are reduced to straight lines

Retail price in lire per kilogram in Italy
(1941 - 1953)

Year	Bread	Rice	Potatoes
1941	2.21	2.27	1.18
1942	2.24	2.30	1.61
1943	2.40	2.49	2.15
1944	12.34	34.85	19.53
1945	24.24	78.00	33.00
1946	37.00	120.00	39.00
1947	73.00	232.00	50.00
1948	93.00	152.00	34.00
1949	99.00	122.00	32.00
1950	96.00	115.00	47.00
1951	96.00	120.00	44.00
1952	102.00	126.00	44.00
1953	107.00	130.00	48.00

with angular coefficient a . Pareto made use of this system to represent the distribution of tax-payers to family tax in 23 Italian towns in 1887, classified according to the amount of their income, and similar distributions of income receivers in other countries.

It should be observed that a logarithmic chart has no zero line and cannot be used if the phenomenon has a *nil* value (or negative values) at any time.

After Cartesian co-ordinate diagrams, *polar co-ordinate diagrams* are particularly important, and are used to represent phenomena having a cyclical character.

For their construction, a fixed point o , called the *pole*, is taken on a straight line called the *polar axis*, and the points representing the phenomenon are then plotted by means of segments, called *radius vectors*, joining them thereto, and of angles made by the segments with the polar axis, called *polar angles*.

The radius vector and the polar angle, which determine the position of any point on the plane, are the two *polar co-ordinates* (fig. 18). The variable measurement is indicated by the size of the polar angle, while the value of the function is indicated by the length of the radius vector. For every cycle of the phenomenon the variable is represented by an angle of 360 degrees, so that the corresponding phases of the various cycles of the

phenomenon fall on the same radius vector and are therefore easily comparable.

If the extreme points of consecutive radius vectors are joined, a *polar curve* is obtained and the diagram is called a *joined polar co-ordinate diagram*. This curve can be formed of segments of straight lines or better of arcs if Archimedes' spiral, as is especially advised in the case of continuous phenomena, which is in accordance with the hypothesis that, from one term to the next, variation takes place proportionally to the interval elapsed.

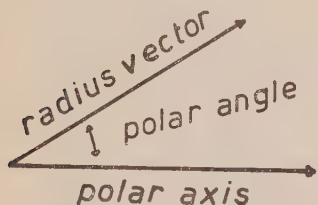


Fig. 18 — Elements of the polar coordinate diagram

An example of this type of representation is given by figures 19, (vectors joined by segments of straight lines) and 20 (vectors joined by arcs of Archimedes' spiral) which indicate the average monthly temperature in Rome in the years 1950 and 1951 (*Annuario Statistico italiano* 1952, page 12).

In certain cases it may be useful make use of *connecting line diagrams*. These consist of two parallel straight lines (vertical or horizontal), on each of which, starting from the origin, segments are marked off proportionally to the intensities that a certain quantitative character presents

in two different groups of elements, between which one wishes to measure the resemblance with regard to the character in question; the extremes of the segments which measure the intensities of the character in two corresponding elements are then joined two by two.

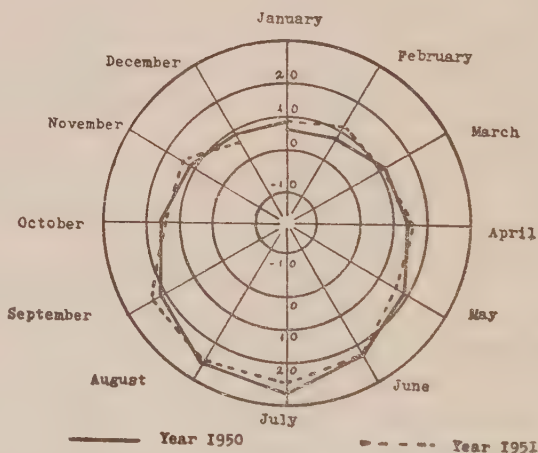


Fig. 19 — Average monthly temperature in Rome (January 1950-December 1951) (vectors joined by segments of straight lines)

If, for example, we wish to demonstrate the relation ship between the productivity per hectare of various crops in Europe and America in 1950, we shall construct figure 21.

The same system can be followed in order to demonstrate the relationship between characters of different generations of individuals, of animals or of plants as a function of given environmental or breeding phenomena, recourse being had to several consecutive parallels on each of which are measured the data of the components of one generation.

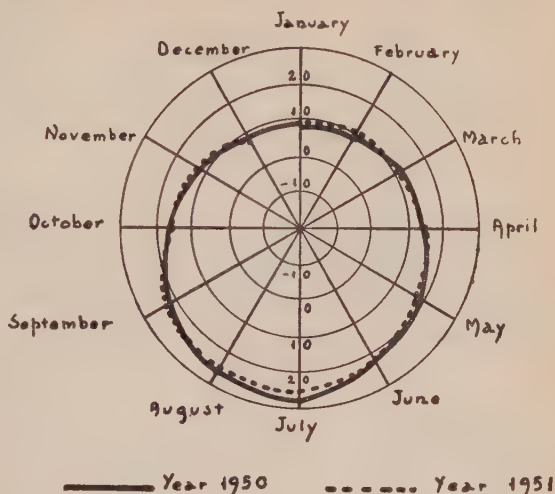


Fig. 20 — Average monthly temperature in Rome (January 1950 to December 1951) [vectors joined by arcs of Archimedes' spiral]

Average monthly temperature in Rome

Month	1950	1951
January	7.1	9.2
February	9.1	10.6
March	11.8	10.7
April	13.7	13.8
May	18.9	17.6
June	22.3	22.3
July	27.1	24.8
August	25.7	25.0
September	21.9	22.9
October	16.8	15.7
November	12.2	13.3
December	9.3	7.9

Let us suppose, for example, that we wish to represent the effect of a given crossing of races on the development of a certain character, on the basis of Mendel's law: we can do so by means of a diagram such as that of figure 22.

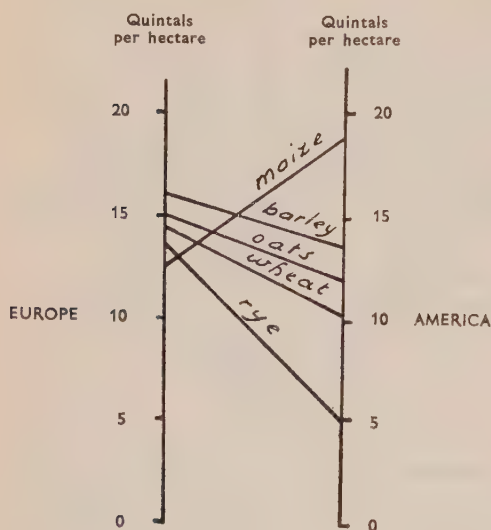


Fig. 21 — Productivity of various crops in Europe and America

B) *A quantitative phenomenon as a function of two qualitative phenomena.*

In order to represent a quantitative phenomenon as a function of two qualitative phenomena (mutables), it is necessary to consider the combinations to which the m possible modalities of the one mutable with the n possible modalities of the other give rise, a total of $m \cdot n$ combinations.

Given the modalities of the first mutable a_1, a_2, \dots, a_m and those of the second b_1, b_2, \dots, b_n , the representative scheme of

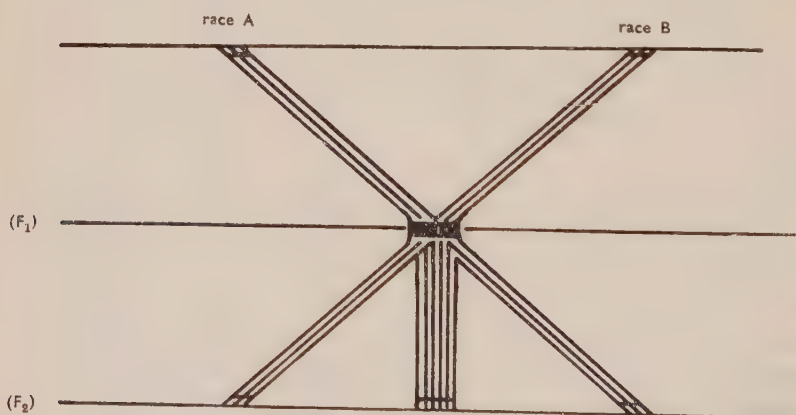


Fig. 22 — Representation of the effects of crossing according to Mendelian laws for first generation (F_1) and second generation (F_2)

the said combinations can equally well assume either of the two following forms :

1st form: $\underbrace{b_1, b_2, \dots, b_n}_{a_1} \quad \underbrace{b_1, b_2, \dots, b_n}_{a_2} \quad \dots \quad \underbrace{b_1, b_2, \dots, b_n}_{a_n}$

2nd form: $\underbrace{a_1, a_2, \dots, a_m}_{b_1} \quad \underbrace{a_1, a_2, \dots, a_m}_{b_2} \quad \dots \quad \underbrace{a_1, a_2, \dots, a_m}_{b_n}$

Corresponding to each of the different combinations, the respective values of the function can be represented by means of geometrical figures, such as those indicated in the paragraph A, in which a linear dimension or an area is proportional to those values.

If, for example, we wish to represent the distribution of the hectares wooded with resinous trees and those with broad-leaved trees in the mountain, hill and plain zones of Italy in 1951, as indicated in the following table, since for the kinds of woods we have two modalities (resinous and broad-leaved trees) and for the heights of the zones three modalities (mountain, hill and plain), the number of combinations will be 2 multiplied by 3, corresponding to 6 modalities.

If we wish to represent the respective areas of the various combinations of kinds of wood and zones by means of segments, we shall raise on a straight line two groups of three ordinates each. The two groups correspond to the two types of the woods ; the ordinates, on the contrary, correspond to the three types of the heights of the zones ; the heights of the ordinates show the wooded

Distribution of the wooded area in Italy
(from the Annuario Statistico Italiano of 1952 page 181)

	Mountain	Hillyzones	Plain	Total
Resinous woods. (in hectares). .	862.194	131.030	46.135	1.039.539
Broad-leaved tree woods. (in hectares).	650.445	354.625	57.093	1.062.163
TOTALS . . .	1.512.639	485.655	103.408	2.101.702

hectares occupied by each type of wood in each zone (fig. 23). We can also construct on a straight line three groups of two or-

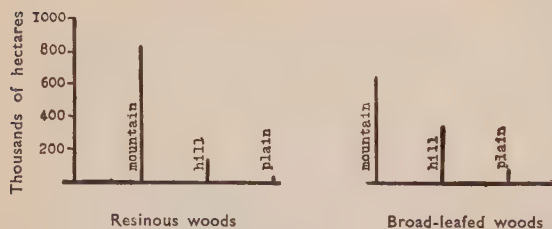


Fig. 23 — Distribution of wooded area in Italy

ordinates each. In this case the three groups correspond to the heights of the zones, and the two ordinates of each group to the resinous and broadleaved tree woods (fig. 24).

The same data can also be represented in other ways, for example, by means of two circles, representing each type of wood and subdivided into sectors of size proportional to the respective zones, as indicated in figure 25, a type of representation known as a *component pie chart*.

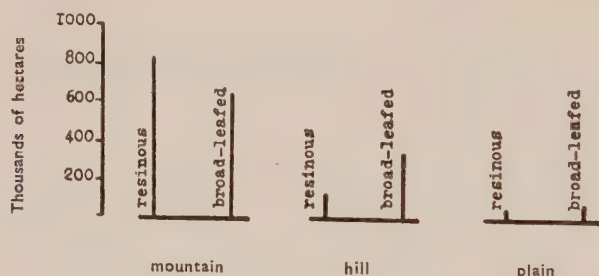


Fig. 24 — Distribution of wooded area in Italy (same data as in fig. 23)

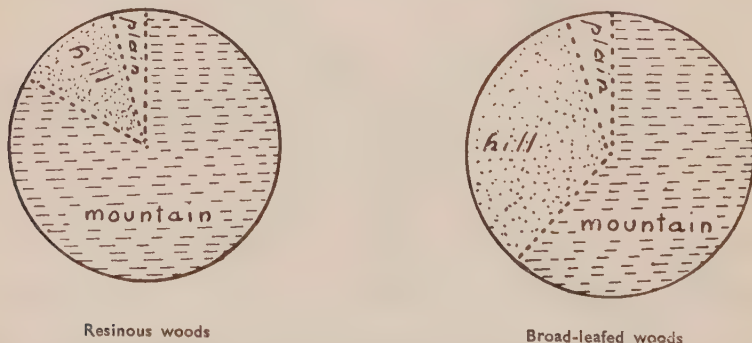


Fig. 25 — Distribution of wooded area in Italy. Component pie chart (same data as in figg. 23 and 24)

One can also construct three circles whose areas are proportional to the quantity of wooded hectares in the three zones, then dividing each circle into sectors proportional to the number

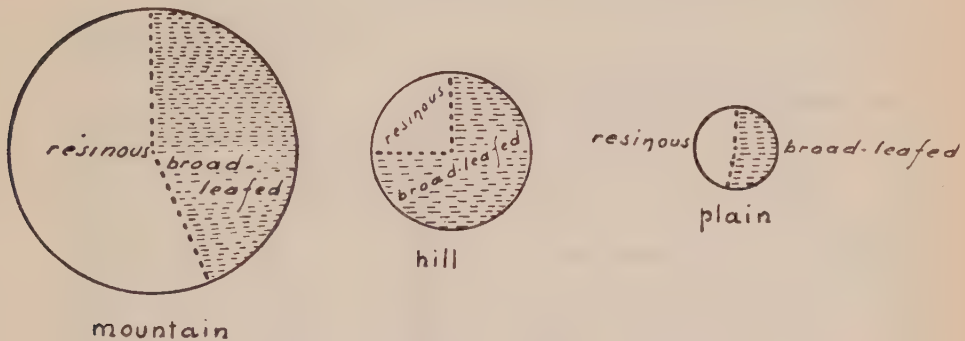


Fig. 26 — Component pie chart Distribution of wooded pie chart (same data as in figg. 23-24-25)

of hectares, in each height zone, wooded with resinous woods and with broad-leaved trees respectively, as in figure 26.

Instead of using circles, one can proceed in the same way with rectangles, which have the advantage that they can be joined together, thus presenting a single surface divided into several parts according to the modalities of the phenomenon, each part being then subdivided according to the modalities of the other phenomenon (figure 27 and 28). This type of representation is called a *hologram*.

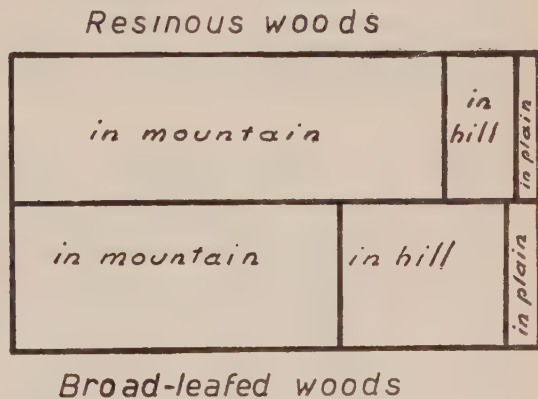


Fig. 27 — Distribution of wooded area in Italy (same data as in figg. 23-26)

B') *A quantitative phenomenon as a function of one qualitative and one quantitative phenomenon.*

When the function depends on one variable and one mutable, we can suppose for a moment that the function depends

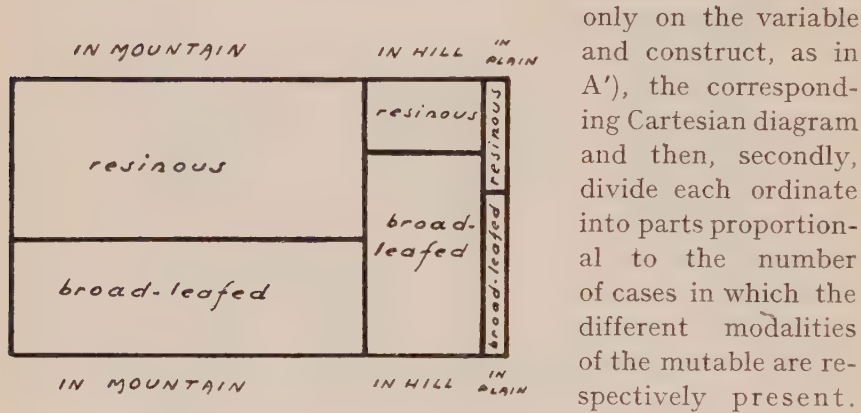


Fig. 28 — Distribution of wooded area in Italy (same data as in figg. 23-27)

only on the variable and construct, as in A'), the corresponding Cartesian diagram and then, secondly, divide each ordinate into parts proportional to the number of cases in which the different modalities of the mutable are respectively present. We then join the corresponding points of division of the consec-

utive ordinates so that the whole area comprised within the first diagram and the *x*-axis is divided into as many zones as there are modalities of the mutable, zones which can be better distinguished by means of different colors or shading.

Italian production of food cereals
(in thousands of quintals)

	1946	1947	1948	1949	1950	1951
Wheat	61.256	47.021	61.657	70.723	77.736	69.618
Maize.	19.089	19.197	22.497	22.110	19.230	27.497
Rice (grain)	4.885	6.355	6.188	6.102	7.064	7.292
TOTALS	85.230	72.573	90.342	98.935	104.030	104.407

An example of this type of representation can be constructed as in figure 29 on the basis of the data given in the preceding table.

Here the process is applied to an ordinate diagram; it can be applied also to an histogram, obtaining a *subdivided histogram*.

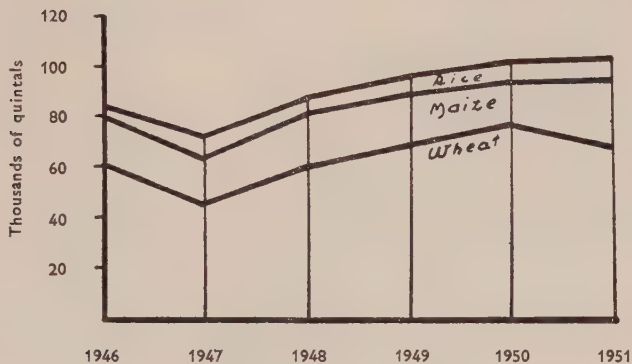


Fig. 29 — Italian production of food cereals

State of the sky over Rome in 1951

Months	Cloudy	Number of days	
		Mixed	Clear
January.	15	12	4
February	12	11	5
March.	11	15	5
April	9	16	5
May.	14	14	3
June	1	21	8
July	1	12	18
August	2	12	17
September.	4	18	8
October	6	18	7
November	5	17	8
December	6	10	15

One can proceed in a similar way for cyclical phenomena on the basis of polar diagrams, when a single cycle or the mean of several cycles is considered, and construct *component polar diagrams*.

If, for example, we wish to indicate the state of the sky over Rome in the various months of the year 1951 on the basis of the data in the following table, we shall have figure 30.

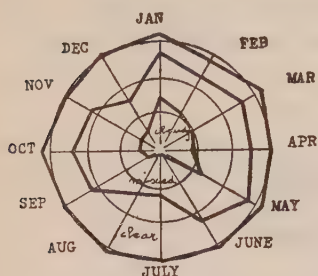


Fig. 30 — State of sky over Rome in 1951

B'') A quantitative phenomenon as a function of two other quantitative phenomena.

When it is necessary to represent a quantitative phenomenon as a function of two other quantitative phenomena, i.e. of two variables, a Cartesian co-ordinate diagram can again be used, but in this case with three axes instead of two. The three dimensional figure (i.e. the figure of a solid) thus obtained is called a *stereogram*.

As an example let us consider the composition according to age of the agricultural population of a given country as shown in consecutive censuses.

We first have recourse to the system of two Cartesian co-ordinates and on this we represent the agricultural population up to 10 years of age in the consecutive censuses; then on another Cartesian diagram of the same scale we construct the graph of the agricultural population from over 10 to 20 years, and other similar successive graphs for the population from over 20 to 30 years, from over 30 to 40 years, from over 40 to 50 years and so on.

These diagrams are then connected to one another by means of a third axis, the *z*-axis, vertical to the first two and on which distances will be taken from the point of origin *o*, proportional to the age-groups comprised in the two-dimensional diagrams previously formed and thus arranged parallel to one another.

This arrangement of the diagrams, reproduced in the plane from an oblique viewpoint in accordance with the laws of perspective, gives a three-dimensional representation, as seen in figure 31.

The use of this type of representation should be limited in view of the complicated construction and of the difficulty of appreciating the behaviour of the phenomenon.

Another three-dimensional system of representation is that of *contour curves* or *isometric curves*, analogous to that widely used in geography and meteorology to represent orographical relief, depths of sea, temperatures, barometric pressures, etc. in various places.

In this type of representation every contour curve connects all the points in which the phenomenon presents the same intensity, and the curves follow one another in accordance with a specific scale, i.e. with a given difference of value between two consecutive curves. According

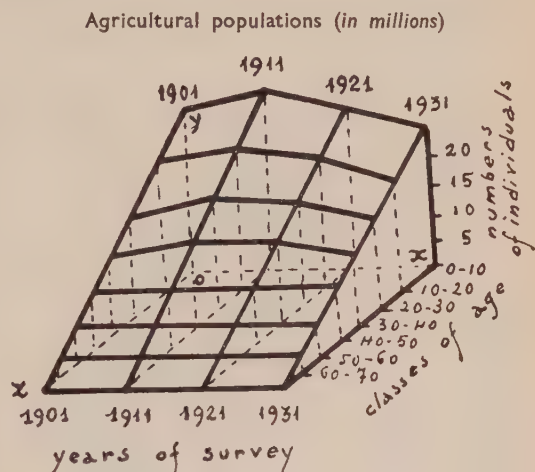


Fig. 31 — Agricultural population according to years of survey and classes of age

to whether the phenomenon increases more or less rapidly, the contour curves are more or less close to one another. (See fig. 32 constructed with the data of the following table).

Marriages celebrated in the Republic of Mexico in the year 1930, classified according to the ages of brides and bridegrooms.

Ages of bridegrooms	Ages of brides								
	Less than 20 years	15-19 years	20-24 years	25-29 years	30-34 years	35-39 years	40-49 years	50-59 years	Over 59 years
Less than 20 years	1444	11162	1248	150	19	10	10	0	0
20-24 "	989	25082	13368	1297	271	76	76	6	0
25-29 "	273	7359	9383	3570	615	187	110	13	1
30-34 "	177	1750	2982	2582	1059	303	113	11	1
35-39 "	29	777	1335	1616	1292	775	222	23	1
40-49 "	28	390	801	991	1104	1213	859	107	6
50-59 "	9	90	197	276	358	427	661	258	26
Over 59 years	1	23	54	89	111	163	366	225	114

A connecting-lines diagram can also be used to represent the relationships between three quantitative phenomena by using three radii with equal angles between them, i.e. angles of 120 degrees. On each of these lines the intensity of one of the phe-

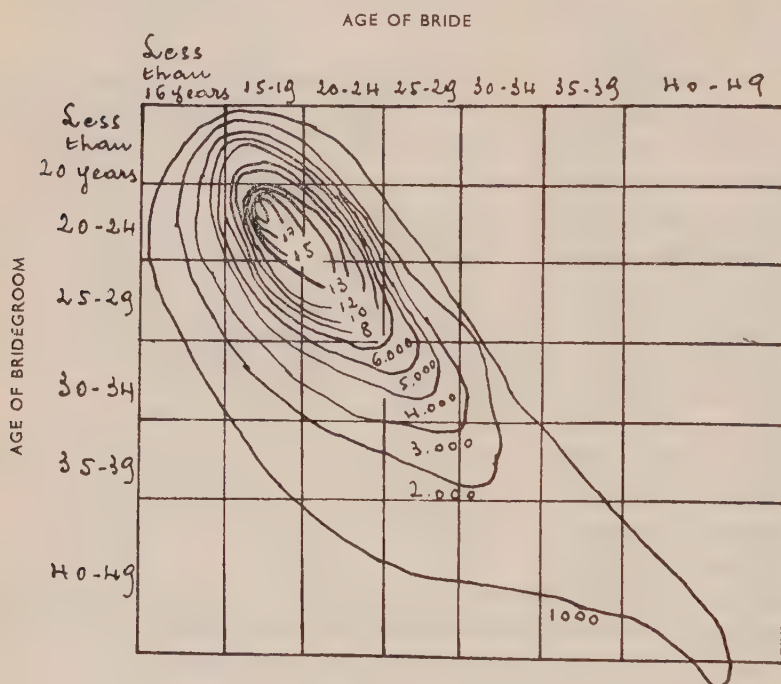


Fig. 32 — Marriages according to the age of bride and bridegroom (Mexico, 1950)

nomena is measured. By means of segments of straight lines one then joins the points which on the three radii represent the corresponding intensities of said phenomena, as shown in figure 33, which indicates the numbers of livestock in Latium and in Tuscany in 1950, classified according to the 3 groups of cattle, horses and pigs, based on the following table.

	Cattle	Horses	Pigs
Lazio	291.569	135.530	203.540
Tuscany	469.898	65.951	325.848

When the sum of the intensities of the three phenomena is constant, one can make use of *triangular co-ordinate diagrams*.

According to a well-known theorem of geometry, if from any point inside an equilateral triangle the perpendiculars to the sides are drawn, the sum of these three perpendiculars is equal to the height of the triangle.

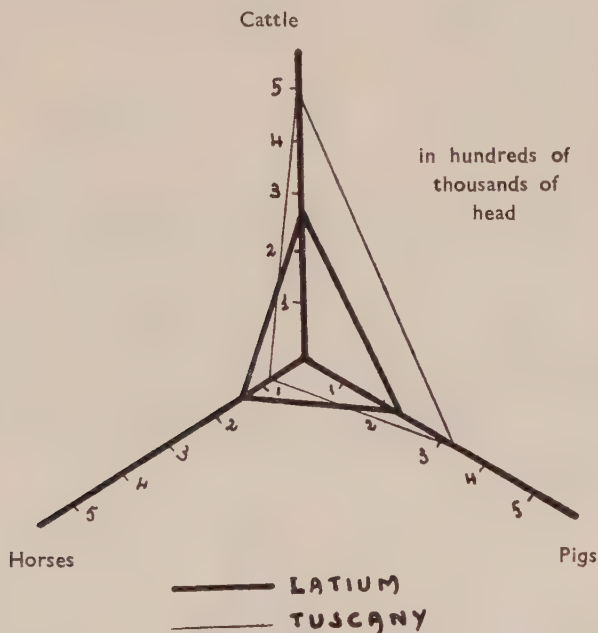


Fig. 33 — Distribution of livestock in three main groups
(Latium and Tuscany, 1900)

If we make the height of the equilateral triangle proportional to the sum of the intensities of the three phenomena, there will always be a point inside the triangle, which will represent any combination of their intensities, and from which the perpendiculars to the sides will be proportional in length to the intensity of each of three phenomena considered.

If, for example, we wish to represent the following percentage composition of the agricultural workers of a given country: tenant-farmers 20 %, crop-sharers 35 %, labourers 45 %, it is sufficient to construct an equilateral triangle of height equal to the total of 100 and to determine the point P which, on the

same scale, has the distance 20 from the side which represents the tenant-farmers, 35 from that which represents the crop-sharers, and 45 from that which represents the labourers, as illustrated in figure 34.

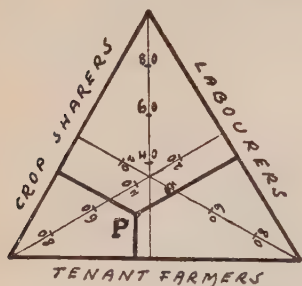


Fig. 34 — Distribution of agricultural workers in three main groups

C) *A quantitative phenomenon as a function of three or more phenomena.*

It is not easy to establish general criteria for the graphical representation of a function which depends on three or more quantitative and qualitative phenomena; the criteria must be established for the individual case.

If, for example, we wish to represent simultaneously the agricultural population by age, sex and civil status, we can use two *juxtaposed component bar charts*, or two *juxta-posed subdivided histograms*, one of which subdivides the males according to civil status and age; and the other similarly subdivides the females, as in shown in figure 35.

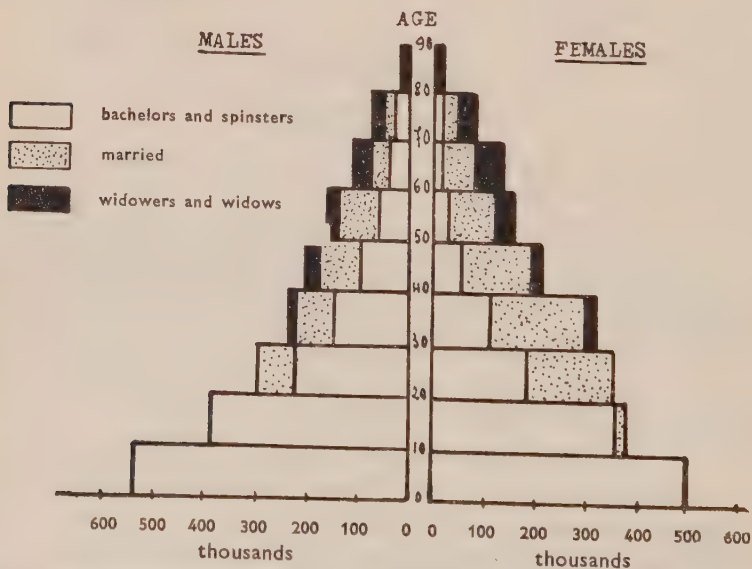


Fig. 35 — Distribution of population according to age, sex and civil status.

When it is a question of comparing four phenomena with one another, recourse can be had to *connecting-lines diagrams with four axes*. These axes must be equi-distant from one another and correspond to those of the tetrahedron reproduced in perspective on the plane. On each of these the intensity of one of the phenomena is measured, and the points representing the corresponding values of the function on the four axes are joined by appropriate connecting-lines.

III. 1. 2. *Cartograms.*

Cartograms are statistical graphical representations based on geographical, chorographical or topographical maps, more or less synthetically reproduced, and which bring out the relationships existing between statistical phenomena and aspects of the geographical environment in a wide sense.

According to the kind of statistical data on which they are based, they can be divided into two categories, the first of which comprises those cases in which the distribution of the phenomenon is indicated *in relation to its geographical location* and the second, comprises those cases in which the intensity of the phenomenon is indicated *in relation to territorial, political or administrative divisions*.

To the first category belong the following types of cartograms : a) those with single or multiple symbols ; b) those with isometric curves ; c) those with band charts.

To the second category belong political-administrative division cartograms, a) with ordinary shading by classes of values, b) with pluri-significant shading, and c) cartodiagrams.

Cartograms for geographical distribution.

A. — *Single or multiple symbol cartograms* are the simplest type of cartograms and consist of a geographical background on which by means of symbols marked in the appropriate localities, the distribution of one or more phenomena is indicated (using different symbols for different phenomena) in the environment in question, such as, for example, factories, warehouses, cheese



Fig. 36 — Multiple symbol cartogram.
Geographical distribution of firms in the O.P.I.C. Groups

dairies, tractor stations, canning factories, slaughter houses, milk centres etc. (fig. 36).

If the phenomenon is of a continuous nature, the frequency of the single symbols, generally dots, can give rise to a continuous area with a greater or lesser intensity of dots according to the intensity of the phenomenon in question (fig. 37).

B. — *Isometric curve cartograms* are used to represent continuous phenomena, such as temperature, altitude, depth, etc., and constitute the application of contour or isometric curves to maps. According to the specific phenomenon which they indicate, the curves take special names, such as *isophysic* curves to indicate



Fig. 37 — Single symbol cartogram (by continuous distribution) of production of grapes (Year 1953)

equal altitude above sea-level, *isobathic* curves for equal depths below the level of the sea, *isothermic* curves for mean temperature (fig. 38), etc.

The more numerous are the isometric curves used, the more detailed is the resulting description of the distribution of the phenomenon, but beyond a certain limit too large a number of curves creates confusion. The determination of the most suitable limit is left to the intuition of the compiler of the car-



TEMPERATURE IN CENTIGRADE DEGREES AT SEA
LEVEL

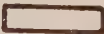



LESS THAN 6°  FROM 6° TO 8° 
FROM 8° TO 10°  MORE THAN 10° 

Fig. 38 — Isothermic curves for mean temperature
(Average winter temperature in 1935)

togram, who should take into account the purposes which he has in mind.

C. — *Band charts* are used for the representation of the intensity of traffic in land, river, sea or air channels of communications, or of international trade as well as of movements of persons or capitals. They can be executed in two ways ; in the first meth-



Fig. 39 — Band chart. Intensity of sea traffic

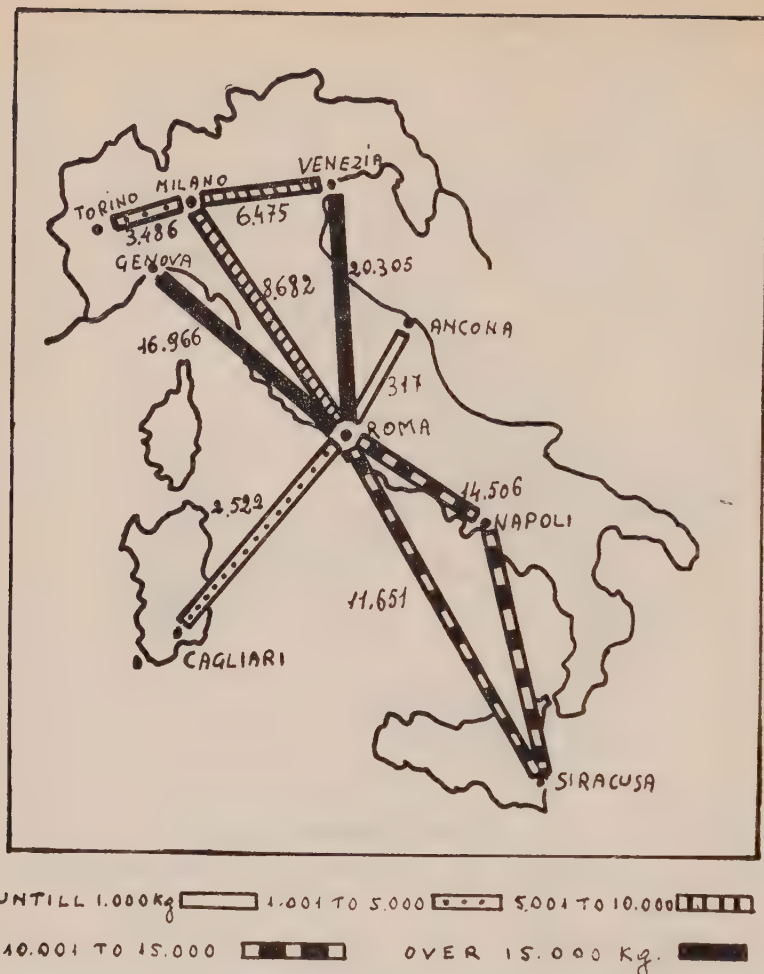


Fig. 40 — Band chart.
Freight transported by the Italian airlines in the year 1935 (data of the Ministero dell'Aeronautica)

od, on a cartographical background the line which represents the channel of communication is drawn the thicker, the greater the volume of traffic which has occurred in that stretch (fig. 39); and in the second method, the entire length of the line of communication is marked by a band of equal dimension in all its length, but variously shaded according to the intensity of the phenomenon (fig. 40). This second method is preferable when the former covers the cartographical back-ground too much.

Cartograms for administrative

It is often necessary to represent geographical distributions of statistical phenomena from data supplied by government and administrative offices in general which give only overall figures or mean values for the entire area covered by the various political-administrative divisions, such as regions, provinces, communes, without any possibility of identifying the values for the various localities composing these territorial divisions. In these cases nothing can be done except to indicate the overall or mean intensities generically for the entire area of these divisions, which is done in the following three ways.

A) Cartograms with simple coloration of shading.

The range of variation of the phenomenon is divided into classes of values, to each of which corresponds a different intensity of coloration or shading, the scale of coloration or shading adopted being indicated in the margin of the cartogram (fig. 41).

Sometimes it is useful to determine the mean intensity of the quantitative phenomenon and to divide the intervals of variation above and below the mean into an equal number of partial intervals, using one coloration for the intensities above the mean and a different coloration for those below the mean, both colorations increasing in intensity as the distance from the mean increases.

Sometimes, on the contrary, if the range of variation is very wide and the distribution of the values strongly symmetrical, the two intervals of variations above and below the mean should each be divided into parts proportional to the terms of a geometrical progression of ratio greater than unity, starting from the extreme intensity which is least distant from the mean, and proceeding in such a way that this separates an equal number of intervals above and below the mean.

We adopted these principles in the construction of the *Atlante Statistico Italiano* (Central Institute of Statistics), in which mean natality and mortality in the five-year period 1910-1914 (before the first world war) and in the three year period 1922-24 (after the first world war) as well as density of population according to the census of 1931 were represented. For

natality and mortality we used arithmetic scales of ten degrees (five above and five below the respective mean intensities of natality and of mortality in Italy in the two periods) : for density,



Fig. 41 — Shaded cartogram with increasing intensity.
Wheat production per hectare in Italy in 1953

on the contrary, we adopted a geometrical scale, also of ten degrees (five below and five above the mean density in Italy in 1921).

B) Cartograms with pluri-significant shading.

By means of this system the area representing each territorial division is shaded with parallel broken lines, in which, if three consecutive vertices are considered, the height of the triangle indicates the intensity of one phenomenon, the base indicates the intensity of another phenomenon, while the distance between two consecutive broken lines, measured normally between their parallel sides, is inversely proportional to the inten-

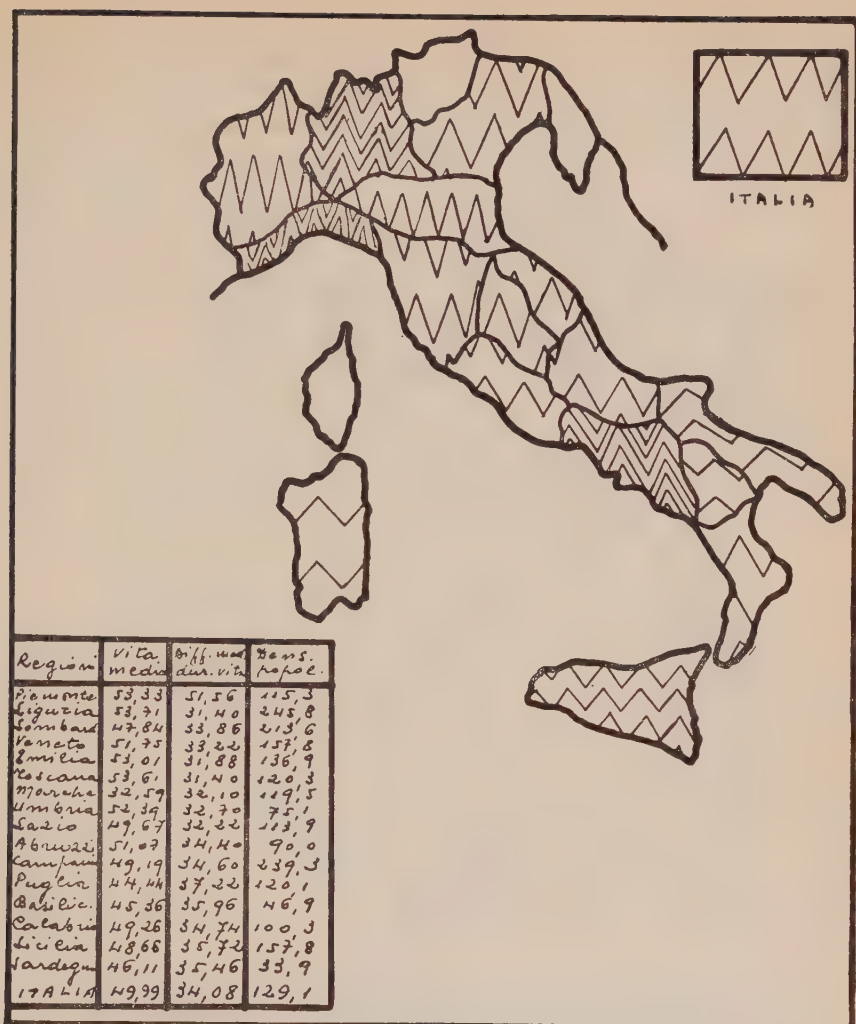


Fig. 42 — Cartogram with pluri-significant shading. Average expectation of life, mean difference of duration of life and density of population in the regions of Italy at 1st December 1921 (from: *Annali di Statistica*, Serie VI, vol. VIII)

sity of a third phenomenon. This system was used by us to represent the average length of life (height of the broken line), the mean difference of duration of life (width of every segment of the broken line), and the density of the population (distance between parallel broken lines) in the various regions of Italy (fig. 42).

C) *Cartodiagrams.*

These are obtained by means of diagrams which are placed in the areas representing the various territorial divisions of the cartographic al background, and indicate the intensity of one or more quantitative phenomena relating to these divisions (fig. 43).



Fig. 43 — Cartodiagram. Cattle in Italy (Year 1953)

As many sub-types of cartodiagrams are thus obtained as there are types of diagrams which can be represented in the areas corresponding to the various divisions.

III. 2. UTILITY AND DRAWBACKS OF GRAPHICAL REPRESENTATIONS

III. 2. 1. *Utility of graphical representations.*

Graphical representations offer many advantages but also present certain drawbacks. Let us begin by examining the former.

1) *They allow the trend of the phenomenon to be assessed at a glance.*

The trend of a phenomenon can also be deduced from a statistical table, but only after a careful and detailed examination and only by people accustomed to statistics ; it can be more clearly perceived when it is expressed by means of a formula, but here too a certain technical knowledge is necessary before a person can deduce correctly therefrom. What cannot be deduced from statistical tables or from formulae without careful examination and without special experience or technical knowledge, can, however, be obtained immediately and by anyone from diagrams. This difference in the speed of perception between graphs and tables or formulae has a physiological and psychological reason which lies in the fact that sight is our most perfect sense and that our mental faculties have developed principally on the basis of that sense.

2) *They serve as a means of verifying the exactitude of the data.*

If, for example, we have statistics of annual international trade which vary with a regular, almost constant movement and we see that the curve which represents the trend of the phenomenon presents a sudden alteration in one year (showing an increase or a decrease), we can presume that there has been either some cause of error in recording the phenomenon, or else think that a new circumstance modifying the movement has arisen, and therefore be induced to verify the exactitude of the data.

It is thus useful to construct, as a check on the calculations, a graph of the results ; this will sometimes allow errors committed by the calculators to be noticed or at least suspected, which otherwise might have passed unobserved.

3) *They are useful in that they sometimes suggest the need for further research.*

There are a number of towns (such, for example, as the town of Augsburg in Germany) of which the population movement during a very long period is known (for this city since the sixteenth century). If we represent the historical series of natality

and mortality of Augsburg by means of diagrams, they show quite considerable oscillations at the beginning, which slowly decrease as we approach the present epoch. It is clear that simply noting this fact does not suffice to explain the phenomenon; it does however suggest the need for seeking the causes. What has been said in relation to variability can equally well

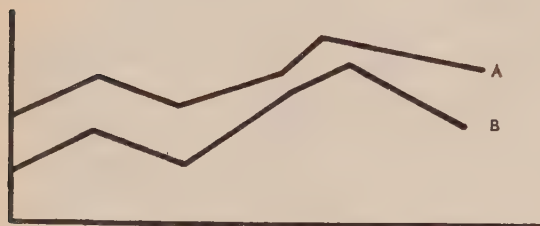


Fig. 44 — Diagram suggesting an association between phenomena A and B

be said for the correlation between two or more phenomena. If we have two numerical series of data corresponding to two phenomena, it will at first sight be difficult to judge whether there is association between the two phe-

nomena, except for cases in which this is very strong. If, on the contrary, we represent the trend of two phenomena graphically, the existence of association between them becomes evident at a glance. The diagram of figure 44, for example immediately suggests an association between the phenomena A and B.

4) *They facilitate the immediate discovery of other relationships between the phenomena without mathematical elaboration (heuristic utility).*

If, for example, we have a diagram such as that in figure 45, in which are represented the variations in the average expectation of life according to age for the inhabitants of Italy, we realise at first sight that the average expectation of life between the ages of 20 and 60 years is in an approximately linear relationship with the increase of age; this

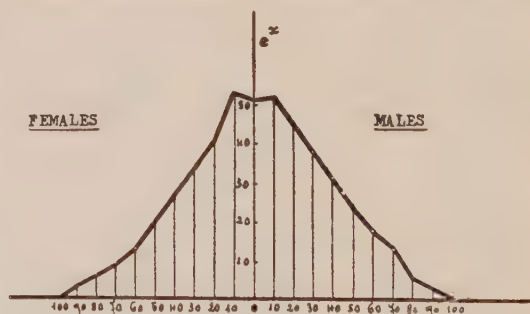


Fig. 45 — Expectation of life according to age and sex in Italy

observation allows us to reach interesting conclusions since, when two phenomena are bound to each other by a linear relationships, the mean value of the one corresponds to the mean value of the other, and it therefore follows that the mean of the expectations of life of a group of females between 20 and 60 years of age or of males between 10 and 60 corresponds to the expectation of life of a person who is of an age equal to the mean age of this group.

Average expectation of life of the Italian population

(1930 - 1932)

Age	Average expectation of life	
	Males	Females
0	53.76	56.00
1	59.71	61.32
2	61.11	62.79
3	60.92	62.62
4	60.37	62.07
5	59.68	61.37
6	58.90	60.60
7	58.07	59.77
8	57.22	58.92
9	56.34	58.04
10	55.46	57.15
20	46.75	48.49
30	38.58	40.41
40	30.39	32.14
50	22.45	23.89
60	15.16	16.13
70	9.05	9.61
80	4.85	5.18
90	2.59	2.82

In connection with human characters we must remember a fact of great importance for statistics, and that is that many characters of the adult man (height, weight, chest measurement, dimensions of the head, face, nose and mouth, length of the various limbs, pulsation, breath frequency, strength of grip and of pushing, speed of movements etc.) stand in what can be considered practically linear relationship with one another. Varia-

tions of stature stand in almost constant proportion with variations of weight, those of weight with those of the chest measurement etc. Analogous results are found for all human characters hitherto examined as well as for many characters of other animal and vegetable species. This can easily be noticed because, when represented graphically, these relationships give rise to a line which, apart from accidental deviations, may be considered a straight line; the deviations are so small that they can be considered accidental and negligible. The admission that human somatic characters are in linear relationship with one another implies the consequence that the mean values of such characters correspond to one another. Therefore a man of average stature, will also be of average weight, average volume of the head and chest, average length of limbs etc.; this is not necessarily true for the single individual, but in the average of cases. On the correspondance of the mean values of human characters is based the *theory of the average man*, of which we shall have occasion to speak further on.

5) *By means of diagrams we can judge the greater or lesser approximation of the results obtained by means of mathematical elaborations.* This is particularly true in the case of interpolation. In fact, the effective curve will approximate more or less closely to the theoretical interpolated curve; the greater or lesser discordance between the two curves will allow us to judge the greater or lesser approximation of the formula of interpolation used. This can be repeated, in general, for all comparisons between effective data and the data calculated on the basis of theoretical considerations.

6) *Diagrams also serve to give us indications on the law of dependence which links the function to the variable*, indications by which we may be led to determine the analytical form of the said law. In fact (for the sake of simplicity we refer to the case of a function depending on a single variable) if the diagram takes the form of a segment of a straight line, of an arc of a circumference, of a parabola, of a hyperbola or of any other curve of which analytical geometry teaches us how to determine the equation, i.e. the analytical nexus between the co-ordinates of all its single

points, this same equation will constitute the analytical expression of the law of dependence of the function in relation to the variable. This has not only and not so much a theoretical value as a practical one, since it allows us to find the intermediate values of the function which have not been actually observed.

When it is a question, as in the concrete cases that appear in Statistics, of diagrams consisting of a discontinuous complex of points, it will naturally be necessary to see if the said points are situated or have a tendency to arrange themselves according to a known type of curve, the equation of which can be determined by processes of *analytical interpolation*, while for an approximate appreciation of the equation it may be sufficient to trace a curve by hand or with the help of an appropriate instrument such as a flexible metallic ribbon, a planimeter etc. (*graphical interpolation*), as we shall see in a later chapter.

7) *They facilitate the comparison between more than one series.* In fact the juxtaposing or superposing of curves relating to more than one series allows us to make a rapid survey of their similarities and differences. The comparison appears even more clearly when each series is represented by means of a point, as we have seen in the case of triangular co-ordinate representations.

III. 2. 2. Drawbacks of graphical representations.

Having examined the advantages of graphical representations, we now have to examine their *drawbacks*.

1) A serious drawback arises from the arbitrary nature of the ratio between the magnitude or scale of measurement used for the variable (or the mutable) and that used for the function.

Let us suppose that we have to plot the variations in the production of wheat according to the years ; if we take abscissae with very small intervals and ordinates with very wide intervals,

the oscillations of the phenomenon may appear very intense (fig. 46). If, on the contrary, the intervals of the abscissae were very wide and those of the ordinates very small, we should have

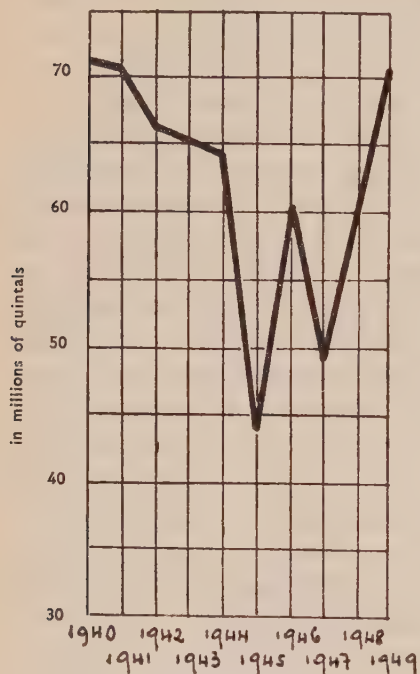


Fig. 46 — Yearly production of wheat in Italy (1940-49)

the impression that the oscillations were much smaller (fig. 47). This difference between the two curves which represent the trend of the same phenomenon, notwithstanding the accuracy of the construction of the two diagrams, depends on the fact that the ratio between the magnitude used for the unit of measurement of the variable and the magnitude used for the unit of measurement of the function is different in the two diagrams.

If we wish to represent the absolute variations of a phenomenon, it will therefore be advisable always to adopt the same scale of natural numbers; if, however, we wish to represent relative variations, it will be necessary to adopt the logarithmic

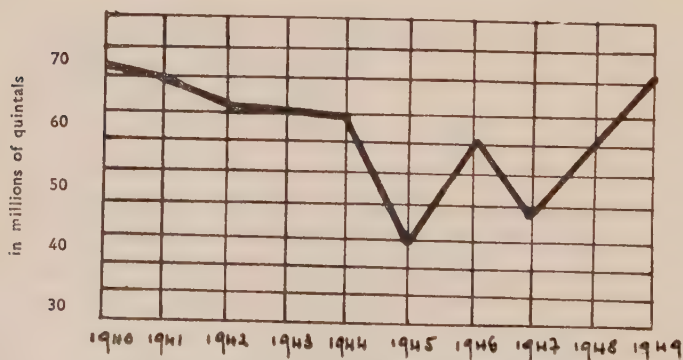


Fig. 47 — Yearly production of wheat in Italy (1940-49)
[same data as represented in fig. 46]

Production of wheat in Italy from 1940 to 1949

Year	Wheat Production (in million of quintals)	Year	Wheat Production (in million of quintals)
1940	71.043	1945	41.766
1941	70.702	1946	61.256
1942	65.754	1947	47.016
1943	65.103	1948	61.653
1944	64.509	1949	70.723

scale, as has already been said, or insure that the scales of natural numbers are such that the means of the various phenomena correspond to ordinates of the same height.

If, for example, we wish to represent graphically the variation of the mortality of the Italian male population from 1931

Year	Deaths due to tubercular illnesses complex	Death due to meningitic tuberculosis and to that of the central nervous system
1931	21,465	2,364
1932	20,877	2,353
1933	20,256	2,003
1934	18,928	1,995
1935	18,573	2,013
1936	18,494	1,973
1937	18,453	1,988
1938	17,484	1,858
1939	16,952	1,718

to 1939 due to the complex of tubercular illnesses and that of the mortality due to meningitic tuberculosis and to tuberculosis of the central nervous system (the figures of which are shown in the following table), and we adopt the same scale for both phenomena, we shall arrive, as is seen from figure 48, at the

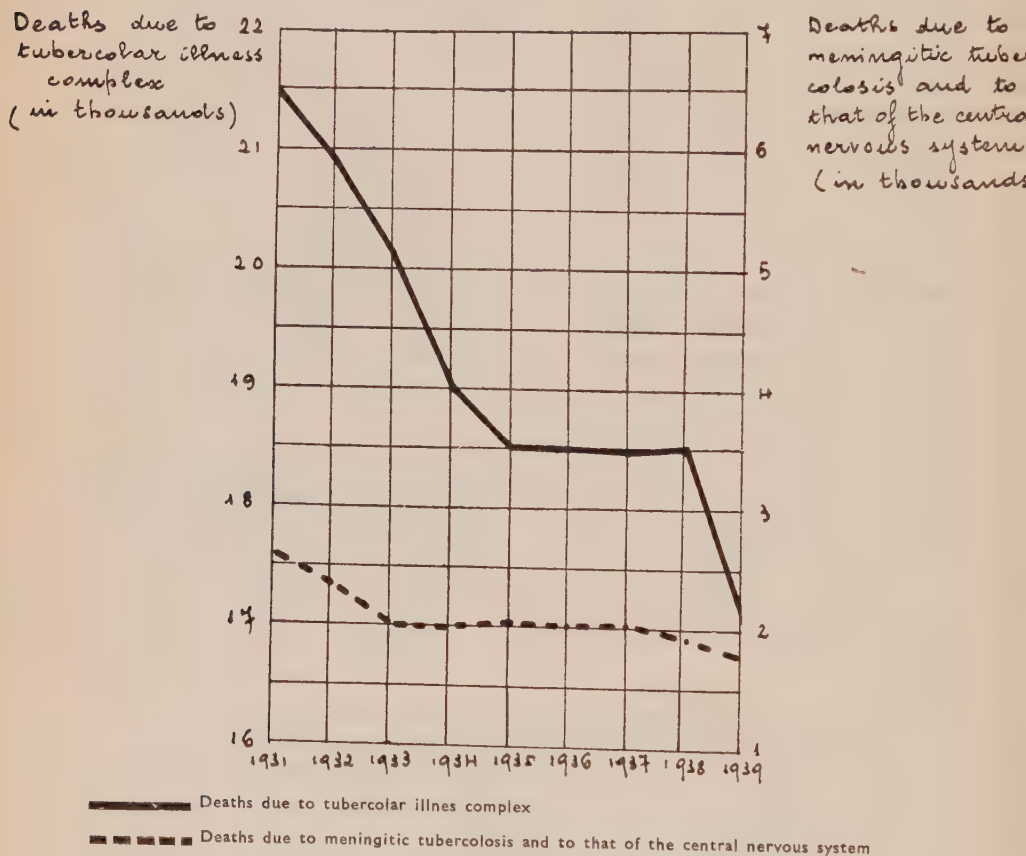


Fig. 48 — Deaths due to various causes (in 1000)

conclusion that mortality for the tubercular illnesses complex has diminished much more rapidly than that for meningitic tuberculosis and for tuberculosis of the central nervous system. This is doubtless true if we consider the phenomenon in an absolute sense, but it is, on the contrary, false if we consider

it relatively. In fact, if we make the mean of each phenomenon equal to one hundred, thus obtaining the following table :

Deaths expressed as percentage of the respective means.

Year	Deaths from the tubercular illness complex	Deaths from meningitic tuberculosis and from that of the central nervous system
1931	112,66	127,89
1932	109,57	114,25
1933	106,31	97,26
1934	99,34	96,87
1935	97,48	97,74
1936	97,06	95,80
1937	96,85	96,53
1938	91,76	90,21
1939	88,97	83,42

we shall find that the mortality for the tubercular illnesses complex has decreased but that the mortality for meningitic tuberculosis and for that of the central nervous system has decreased more rapidly (fig. 49).

2) Another drawback of Cartesian co-ordinate diagrams is that when two or more curves are represented in the same diagram, our eye appreciates the differences between superposed curves not on the basis of the difference between the ordinates, which are proportionate to the values assumed by the two functions corresponding to a determined value of the variable, but on the basis of the minimum distance between the curves, which, when the latter are considerably inclined, can differ substantially from the difference between the ordinates. A clear idea of this drawback is given by the two diagrams of figures 50^a and 51^a. In the first the difference between ordinates remains constant, but without doubt one has the impression that the divergence between the two curves is more marked in proximity to the maximum and minimum points than in the ascending and descending sections. In the second, on the contrary, one has the impression that the divergence between the curves remains constant while the differences between the ordinates vary considerably, being

much greater in the ascending and descending sections than in the proximity of the maximum and minimum points. In order to avoid such an erroneous interpretation, I have proposed that the representation of the curves should be accompanied

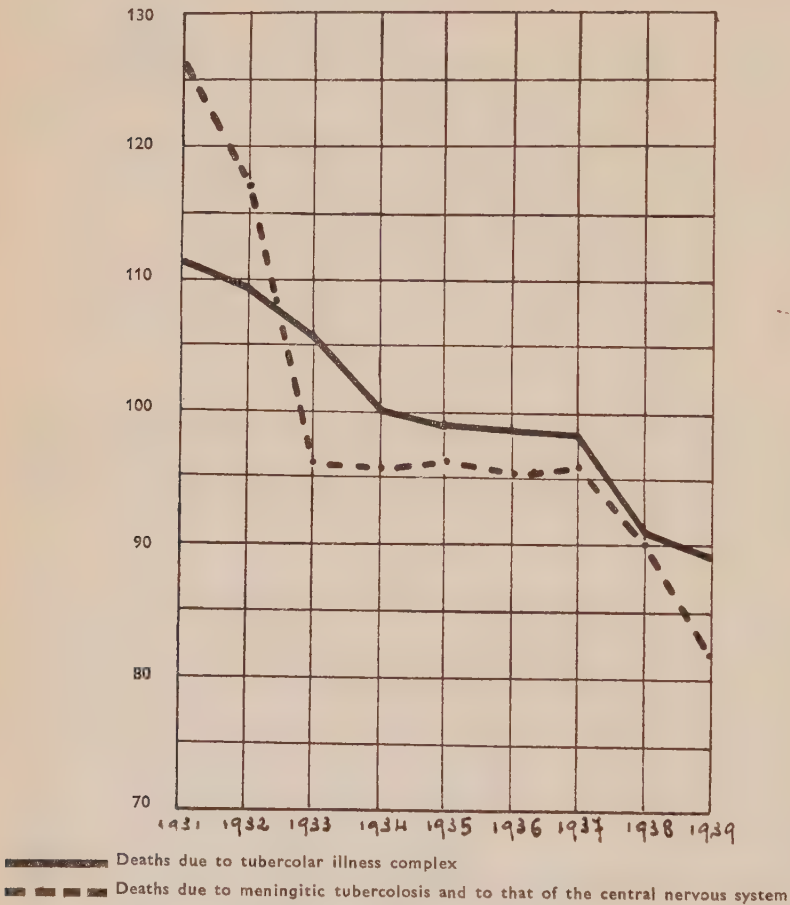


Fig. 49 — Deaths due to various causes (in 1000) (same data as in fig. 48)

by that of their differences, as is done in the diagrams of figures 50b and 51b. Thus, while the normal representation allows us to judge the general trend of the curves, the additional representation enables us correctly to judge their differences at successive points.

3) A number of authors have introduced, among the methods of graphical representation, that of profiles. This method is generally applied to the representation of unconnected series, the terms of which, as is known, do not present a natural order of succession. Substantially it consists in the following: on one of the co-ordinate axes are plotted at equal intervals the points corresponding to the various modalities of the series, and, perpendicularly to these, segments are raised proportional to the intensities (or frequencies) corresponding to those modalities;

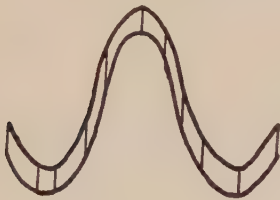


Fig. 50 a



Fig. 51 a



Fig. 50 b



Fig. 51 b

the extremes of these segments are then joined by segments, which together form a broken line, known as a "profile". The graph reproduced here (fig. 52) is taken from an article by Rossolimo.

Subsequently various authors have modified the criteria for selecting the scale of measurement to be adopted in order to obtain heights of segments proportional to the intensities of the single modalities, but have continued to join the extremes of these segments.

Now, this joining — which gives the method its name and therefore suggests that it is the characteristic of the graph from which at a glance the characteristics of the phenomenon can be drawn — can lead to error.

This may be realised immediately, if we observe that, when the order of succession of the modalities is in part or in whole

arbitrary, and can therefore be modified at will, the joined ordinate diagram (and, in particular, the graphical profile) varies

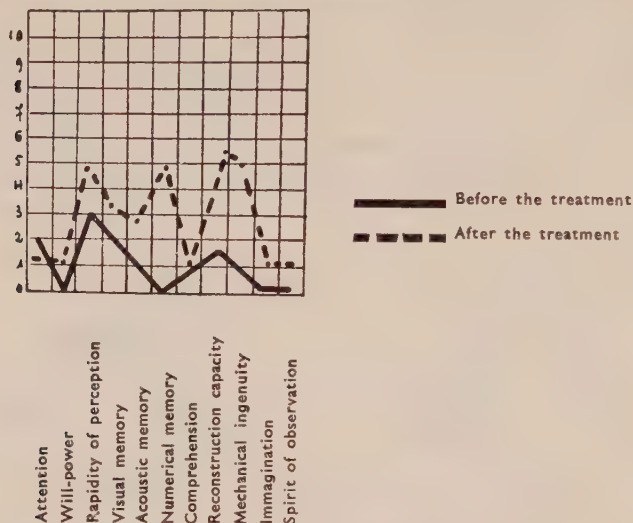


Fig. 52 — Method of profiles. Representation of several mental characteristics of a patient before and after the treatment

according to that order and can leave a radically different impression. Such in the case, for example, for the two following

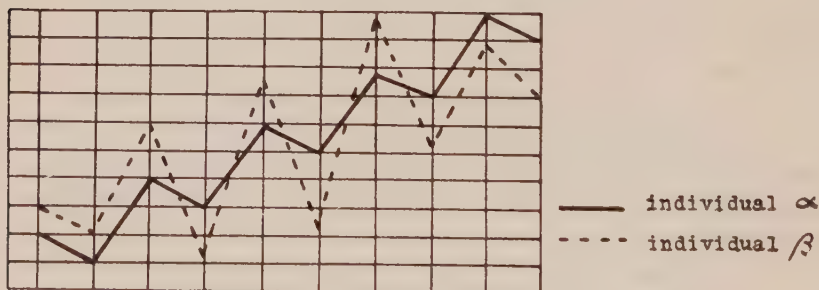


Fig. 53 — Diagram traced with the method of profiles

graphs (figs. 53 and 54) which are, in fact, constructed with the same data given in the following table, but differently ordered.

In the case of series with modalities which do not present a natural order of succession, it is then advisable to make separate graphical representations and for each of these a diagram

Individuals	Characteristics									
	a	b	c	d	e	f	g	h	i	l
α	2	1	4	3	6	5	8	7	10	9
β	3	2	6	1	8	2	10	5	9	7

with a circular or polygonal area, so that the image of the profile ceases to exist and the eye is not led to construct it ideally. A correct graph to substitute Rossolimo's might be that shown in figure 55.

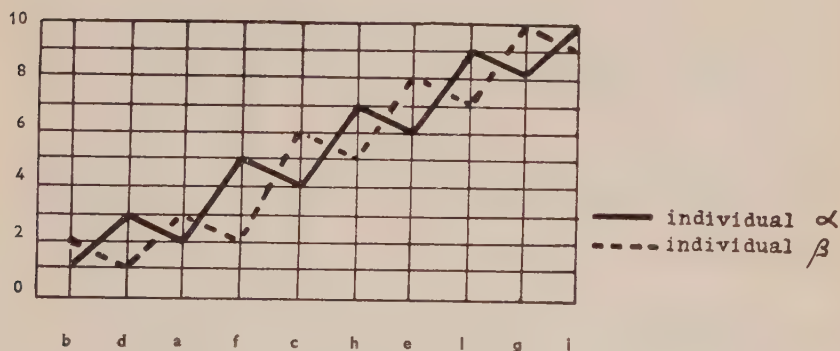


Fig. 54 — Diagram traced with the method of profiles (same data as represented in fig. 53)

4) A drawback of colored graphical representations may arise from the different intensities of the colors used and, with the same intensity, from the different measure with which the different colors impress the eye. This leads to deceptions. One such was put forward by the Hungarian delegation to the committee which had to carry out the partition of Transylvania between Hungary and Rumania. The Hungarian and the

Rumanian communities were represented by means of a cartodia-gram in which every community was represented by a circle of area proportional to the number of its components, a bright red circle for the Hungarian and a straw yellow one for the

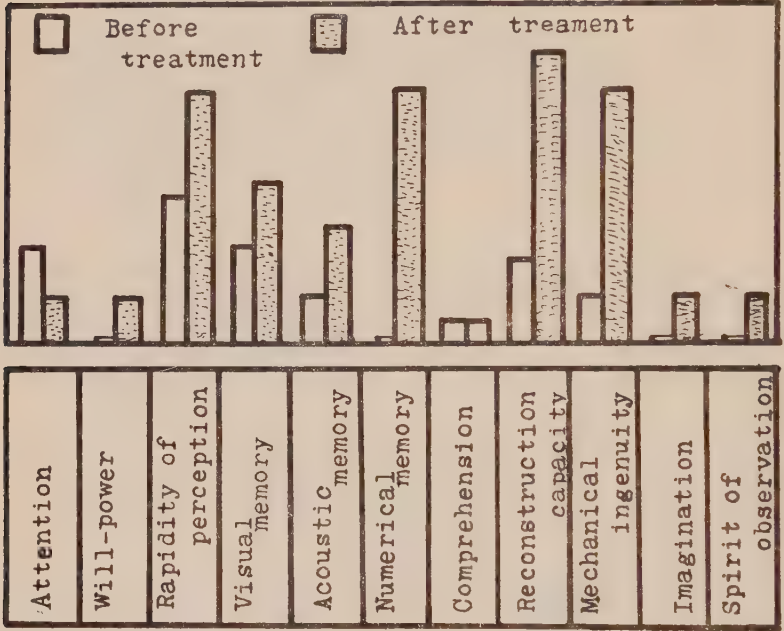


Fig. 55 — A correct representation of the same data represented in fig. 52

Rumanian communities. Despite the fact that the Rumanians were about double the number of the Hungarians, the greater brightness of the red left the impression that the Hungarians constituted the majority.

Publicazioni ricevute. Publications reçues. Publications received. Erhaltene Veröffentlichungen. Publicaciones recibidas.

Periodici. Périodiques. Periodicals. Zeitschriften. Periodicos

Annali della Facoltà di Economia e Commercio, Università di Bari. — Anno XXIX (1958) e XXX (1959).

Annali della Facoltà di Economia e Commercio, Università di Palermo. — Anno XIII (1959).

Arco. — Roma, 1959.

Assicurazioni. — Roma, INA, 1959.

Atti e Memorie dell'Accademia Patavina di Scienze Lettere ed Arti. — Padova, vol. LXIX (1957-58).

Atti Istituto Veneto di Scienze Lettere ed Arti. — Venezia, Tomo CVII, 1958-59.

Banca Nazionale del Lavoro Quarterly Review. — Roma, 1959.

Banca dei Regolamenti Internazionali. — Basilea, Ventinovesima Relazione Annuale, 1° Aprile 1958 - 31 Marzo 1959.

Bancaria. — Roma, 1959.

Bilancio. — Firenze, 1956, 1957, 1958, 1959.

Bollettino della Banca d'Italia. — Roma, 1959.

Bollettino della Federazione Nazionale della Stampa Italiana e della C.U. per l'albo dei giornalisti. — 1959.

Bollettino Mensile di Statistica. — 1959 (e suppl.).

Bollettino della Società Italiana per l'Organizzazione Internazionale. — Roma, 1959.

Bollettino Statistico (Comune di Roma). — 1959.

Bollettino Statistico (Ist. Naz. Ass. Malattie). — Roma, 1959.

Comunicazioni alla stampa. — Milano, 1959.

- Comunità Europea del Carbone e dell'Acciaio** (Notiziario). — Roma, 1959.
- Congiuntura Economica**. — Milano, 1959.
- Conoscersi** (Rivista dell'Associazione Italiana rapporti culturali con la Polonia). — Roma, 1959.
- Endeavour**. — London (edizione italiana), 1959.
- Esso rivista**. — Roma, 1959.
- (II) Giornale dell'impiegato agricolo**. — Roma, 1959.
- Giornale dell'Istituto Italiano degli Attuari**. — Roma, 1958.
- Giornale dell'Università**. — Roma, 1959.
- Idee e Sistemi**. — Roma, 1959.
- L'Industria**. — Milano, 1959.
- Lotta contro la tubercolosi**. — Roma, 1959.
- Mondo Economico**. — Milano, 1959.
- Moneta e credito**. — Roma, 1959.
- Note econometriche**. — Torino, 1959.
- Notiziario ISTAT**. — Roma, 1959.
- Notiziario ortofrutticolo**. — Roma, 1959.
- Notiziario statistico (I.N.A.I.L.)**. — Roma, 1959.
- Notiziario statistico mensile del Comune di Roma**. — Roma, 1959.
- Notizie e Programmi dell'Agenzia Europea di Produttività**. — Parigi, 1959.
- Politica Sindacale**. — Roma, 1958, 1959.
- (II) Politico**. — Pavia, 1959.
- Previdenza Sociale**. — Roma, 1959.
- I Problemi della Sicurezza Sociale**. — Roma, 1959.
- Progresso agricolo**. — Bologna, 1959.
- Quaderni della Società generale immobiliare**. — Roma, 1956, 1957, 1958, 1959.
- Rassegna economica del Banco di Napoli**. — Napoli, 1957, 1958, 1959.
- Rassegna economica Società per azioni**. — Roma, 1959.
- Rassegna di Statistiche del Lavoro**. — Roma, 1959.
- Rendiconti dell'Istituto Lombardo di Scienze e Lettere**. — Milano, 1958.
- Rivista Bancaria**. — Milano, 1959.
- Rivista di Politica Economica**. — Roma, 1959.
- Sintesi Economica**. — Roma, 1959.
- Lo Spettacolo**. — Roma, 1959.
- Statistica**. — Bologna, 1959.
- Statistica del Commercio con l'estero**. — Roma, 1958, 1959.
- Statistiche del lavoro**. — Roma, 1958.
- Stato Sociale**. — Roma, 1959.
- Studi Economici**. — Napoli, 1959.

- Studi e Statistiche.** — Milano, 1959.
- Supplemento al Bollettino Statistiche del Lavoro.** — Roma, 1959.
- Tecnica ed Organizzazione.** — Torino, 1959.
- UTET.** — Torino, 1957, 1958, 1959.
- Bulletin Commission Internationale des Juristes.** — La Haye, 1958, 1959.
- Bulletin Européen.** — Roma, 1959.
- Bulletin Mensuel Economie et Statistique Agricoles.** — Roma, 1959.
- Bulletin Trimestriel des Statistiques du Logement et de la Construction pour l'Europe.**
— Ginevra, 1959.
- (Les) Cahiers français.** — Paris, 1959.
- (Le) Contrat Social.** — Paris, 1959.
- Documentation économique.** — Paris, 1959.
- Eskimo.** — Churchill (Manitoba), 1956, 1957, 1958, 1959.
- Etudes et Conjonctures.** — Paris, 1959.
- (Les) Etudes sociales.** — Paris, 1959.
- Information.** — Paris, 1959.
- Population.** — Paris, 1959.
- Revue de l'Institut International de Statistique.** — La Haye, 1957, 1958.
- Revue de la Mesure de la Productivité.** — Paris, 1959.
- (The) American Statistician.** — Washington, 1959.
- (The) Annals of Mathematical Statistics.** — Ann Arbor, 1959.
- Bank of Finland.** — Helsingfors, 1959.
- Bulletin Bank of Norway.** — Oslo, 1959.
- Bulletin of the Statistical Society of New South Wales.** — Sidney, 1954, 1955,
1956, 1957, 1958.
- Economic Bulletin for Europe.** — Ginevra, 1959.
- Educational Research Bulletin.** — Ohio, 1959.
- The Freeman.** — New York, 1959.
- Index.** — Stockholm, 1959.
- Industrial Bulletin.** — Albany, 1959.
- Journal of the American Statistical Association.** — New York, 1959.
- Journal of the Royal Statistical Society.** — London, 1959.
- Migration News.** — Ginevra, 1959.
- Monthly Vital Statistics Report.** — Washington, 1959.
- Morbidity and Mortality.** — Washington, 1959.
- Newsletter.** — Cambridge (Mass.), 1959.
- Population Bulletin** — 1959.
- Sankhya.** — Calcutta, 1959.
- Statistical Bulletin.** — New York, 1959.

- Statlab Terview.** — Jowa, 1958.
- Think.** — New York, 1959.
- Vital Statistics Special Report.** — Washington, 1959.
- Raumforschung und Raumordnung.** — Köln-Berlin, 1959.
- Schweizerische Zeitschrift für Volkswirtschaft und Statistik.** — Berna, 1959.
- Boletin Banco Central de Reserva del Perú.** — Lima, 1959.
- Boletin de Epidemiologia y Estadistica. Departamento de Estadísticas Sanitarias.**
— Madrid, 1959.
- Boletin de Estadistica.** — Madrid, 1959.
- Boletin de Estadistica.** — Caracas, 1959.
- Estadistica Panamena.** — Estrella de Panama, 1959.
- Resumen estadístico.** — Madrid, 1959.
- Revista Internacional de Sociologia.** — Mexico, 1958, 1959.
- Revista Nacional de Cultura.** — Caracas, 1959.
- Revista Shell.** — Caracas, 1959.
- Trabajos de Estadistica.** — Madrid, 1959.
- Tricolor.** — Caracas, 1959.
- Boletim estatístico.** (Ist. Bras. de Geogr. e Estat.). — Rio de Janeiro, 1959.
- Revista brasileira de Estatística.** — Rio de Janeiro, 1959.
- Revista de Centro de Estudios Demograficos Economicos.** — Lisboa, 1959.
- Revista de la Facultad de Ciencias Economicas y de Administracion.** — Montevideo,
1954, 1955, 1956, 1956, 1958, 1959.
- Statistisk Tidskrift.** — Stockholm, 1959.
- Tilastokatsauksia.** — Helsinki, 1959.
- Ja Ulkomaankauppa.** — Helsinki, 1959.

PUBBLICAZIONI UFFICIALI

- AUSTRALIA (Commonwealth of). — *Census of the Commonwealth of Australia*, 30th June 1954, Vol. VIII *Australia: Part I, Cross-classification of the characteristics of the population. Part II, Cross-classifications of the characteristics of dwellings and of householders. Part III, Population and occupied dwelling in localities* (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Demography* 1957, bull. n. 75 (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Finance, p. II, Commonwealth taxation*, bull. n. 48, 1956-57 (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Labour Report* 1957, bull. n. 45 (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Oversea Trade* 1957-58, bull. n. 55 (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Pocket Compendium of Australian Statistics*, bull. n. 44, Canberra, 1959.
- ID. — *Primary Industries. Part I° Rural industries. Part II° Non-rural industries and value of production* 1956-57, bull. n. 51 (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Transport and Communication* 1956-57, bull. n. 48 (by S. R. Carver, Acting Commonwealth Statistician), Canberra.
- ID. — *Yearbook of the Commonwealth of Australia* 1958, bull. n. 44, Canberra.
- BRASILE. — *Anuario estatístico do Brasil* 1958, año XIX, Rio de Janeiro, dezembro 1959.
- ID. — *Conjuntura economica. Economics and business in Brasil*, August 1959, year VI, n. 8, Rio de Janeiro.
- CINA. — *Report on Fulfilment of the National Economic Plan of the People's Republic of China in 1955. With Statistical Summary*, State Sta-

- tistical Bureau of the People's Republic of China, June 14, 1956, Peking 1956.
- ID. — *Report on National Economic Development and Fulfilment of the State Plan in 1954*, State Statistical Bureau of the People's Republic of China, Sept. 21, 1955, Peking, 1956.
- FINLANDIA. — *Bank of Finland 1958*, Helsinki, 1959.
- ID. — *Foreign Trade 1957-1958*, Helsinki, 1959.
- ID. — *Statistical Yearbook of Finland 1958*, Helsinki.
- ITALIA. — Banco di Sicilia, Istituto di Credito di Diritto Pubblico, *Lineamenti e prospettive di sviluppo dell'economia siciliana*, Palermo.
- COMUNE DI UDINE, Bollettino statistico. — *Annuario statistico 1958*.
- CONFEDERAZIONE GENERALE DELL'INDUSTRIA ITALIANA. — *Annuario 1959*, Roma.
- CONSIGLIO NAZIONALE DELLE RICERCHE. — *Organizzazione generale del C.N.R.* 3^a ediz. aggiornata 1^o marzo 1959, Roma, 1959.
- FINSIDER. — *XXII esercizio sociale 1958-59*, 1959.
- ISTITUTO CENTRALE DI STATISTICA. — *Annuali di Statistica*, anno 88, serie VIII, vol. 10, *Tavole di mortalità della popolazione italiana 1950-53 e 1954-57*, Roma, 1959.
- ID. — *Annuario di Statistica agraria*, 1959.
- ID. — *Annuario di Statistica demografica 1956*, Roma, 1959.
- ID. — *Annuario di Statistica forestale 1958*, Roma, 1959.
- ID. — *Annuario delle Statistiche culturali 1959*, Roma, 1959.
- ID. — *Annuario di Statistiche industriali 1958*, Roma, 1959.
- ID. — *Annuario di Statistiche provinciali 1959*, Roma, 1959.
- ID. — *Annuario di Statistiche sanitarie 1956*, Roma, 1959.
- ID. — *Annuario statistico dell'assistenza e della previdenza sociale 1955-56, 1956-57*, Roma, 1958-59.
- ID. — *Annuario statistico dell'attività edilizia e delle opere pubbliche 1958*, Roma, 1959.
- ID. — *Annuario statistico del commercio interno 1957, 1958*, Roma 1959.
- ID. — *Annuario statistico dell'istruzione italiana 1958*, Roma, 1959.
- ID. — *Annuario statistico italiano 1958*, 1959.
- ID. — *Attrezzature sanitarie degli Istituti di cura 1956*, Roma, 1959.
- ID. — *Bilanci delle amministrazioni regionali, provinciali e comunali. Conti consuntivi*, 1956, Roma, 1958.
- ID. — IX Censimento Generale della Popolazione. III Censimento Generale dell'Industria e Commercio. 4-5 novembre 1951. *Caratteristiche demografiche ed economiche dei grandi comuni.*
Vol. I *Dati riassuntivi e dati dei Comuni non oltre 100.000 abitanti.*, Roma, 1959.

- Vol. II Comuni da 60.000 fino a 100.000 abitanti*, Roma, 1959.
Vol. III Comuni fino a 60.000 abitanti, Roma, 1959.
- ID. — III Censimento Generale dell'Industria e Commercio, 5 novembre 1951.
Vol. XVIII Atti del Censimento, Roma, 1958.
- ID. — *Elezioni amministrative del 1956. Vol. I Elezione dei consigli provinciali, Vol. II Elezione dei consigli comunali*, 1959.
- ID. — *Italian Statistical Abstract*, 1959.
- ID. — Metodi e Norme :
 Serie C n. 2, maggio 1959, *Classificazione delle attività economiche*.
 Serie B n. 5, giugno 1959, *Istruzioni per la rilevazione statistica del movimento della navigazione marittima*.
- ID. — Note e Relazioni :
 n. 5, dicembre 1958, *Ricoverati per tubercolosi, tumori maligni e malattie mentali*.
 n. 6, febbraio 1959, *Indagine sulle scelte scolastiche e professionali degli alunni delle scuole medie inferiori*.
 n. 7, marzo 1959, *Indagine campionaria sulla morbosità della popolazione italiana*.
- ID. — *Popolazione e circoscrizioni amministrative dei Comuni*, Roma, 1959.
- ID. — *Rilevazione nazionale delle forze di lavoro* (Suppl. Annuario Statistiche del Lavoro), 20 ottobre 1958, 20 gennaio 1959, 20 aprile 1959, Roma, 1959.
- ID. — *Statistica annuale del commercio con l'estero 1957*, Roma, 1959.
- ID. — *Statistica degli incidenti stradali 1958*, Roma, 1959.
- ID. — *Statistica della macellazione 1957*, Roma, 1959.
- ID. — *Statistica della navigazione marittima 1957, 1958*, Roma, 1959.
- ID. — *Statistica della pesca e della caccia 1958*, Roma, 1959.
- ISTITUTO NAZIONALE DELLA PREVIDENZA SOCIALE. — *Notizie statistiche 1955-56, 1957*, Roma, 1959.
- ID. — *Statistiche delle pensioni dell'assicurazione obbligatoria. Invalidità vecchiaia e superstiti al 31 dicembre 1958*, Roma, 1959.
- MEDIOBANCA. — *Indici e dati relativi ad investimenti in titoli quotati nelle borse italiane (1948-1959)*.
- MINISTERO DELLE FINANZE. — *Imposte comunali sui consumi dal 1954 al 1956*, Roma, 1958.
- MINISTERO DEI LAVORI PUBBLICI. — *Annuali idrologici 1956 (parte II), 1957 (parte I)*.
- ID. — *Precipitazioni medie mensili ed annue e numero dei giorni piovosi per il trentennio 1921-1950*, n. 24, fasc. VIII, Roma, 1957.
- NEW ZEALAND. — *School Population 1959*.

NORVEGIA. — *Société d'assurances* 1957, Oslo, 1959.

PORTOGALLO

Estatística agrícola 1957, 1958, Instituto Nacional de Estatística, Lisboa, 1958, 1959.

ID. — *(A) Freguesia de Santa Catarina de Lisboa no I quartel do século XVIII (ensaio de demografia historica)*, por M. de Lourdes Akola da Cunha Meira de Carmo Da Silva Ulto, Instituto Nacional de Estatística, Centro de Estudos demograficos, Lisboa, 1959.

ID. — *Gado e Animais de capoeira. Arrolamento geral efectuado em 15 de Dezembro de 1955 no Continente e Ilhas adjacentes*, Instituto Nacional de Estatística, Lisboa, 1958.

ID. — *Inquérito as condições de vida de população da cidade de Coimbra 1953-54*, Instituto Nacional de Estatística, «Estudos», N. 30, Lisboa, 1958.

ID. — *Inquérito Industrial. Distrito de Beja*, Instituto Nacional de Estatística, Lisboa, 1958.

ID. — *Inquérito industrial. Distrito de Castelo Branco* 1957, Instituto Nacional de Estatística, Lisboa, 1959.

ID. — *Inquérito Industrial. Distrito de Portalegre* 1957, Instituto Nacional de Estatística, Lisboa.

ID. — *Inquérito as rendas e a outras características das habitações arrendadas na cidade de Evora em 1955-56*, Instituto Nacional de Estatística, «Estudos» N. 33, Lisboa, 1958.

ID. — *Organização corporativa e Previdência social* 1957, 1958, Instituto Nacional de Estatística, Lisboa, 1958, 1959.

REPUBLICA DOMINICANA

Accidentes del trabajo 1957, Ciudad Trujillo, 1959.

ID. — *Boletin de Divulgaciones estadísticas*, n. 2, febrero 1959.

ID. — *Comercio exterior de la Republica Dominicana*: Vol. VI (1958), Vol. VII (1959), Ciudad Trujillo.

ID. — *Estadística Bancaria*, vol. XXIII (1958).

ID. — *Estadística industrial de la Republica Dominicana* n. 7 (1957), Ciudad Trujillo, 1959.

ID. — *Estadística de los negocios de seguros*, Vol. XXIII (1958), Ciudad Trujillo, 1959.

ID. — *Finanzas municipales, Estadística de Ingresos y Egresos*, Vol. XII (1957), Ciudad Trujillo, 1958.

ID. — *Registro publico. Estadística de operaciones registradas*, Vol. XXII (1957), Ciudad Trujillo, 1958.

ID. — *Sacrificio de ganado*, Vol. XXI (1957), Ciudad Trujillo, 1958.

REPUBLICA DE VENEZUELA

Boletín de comercio exterior, Año I, Mes I, n. 1, 1959.

SPAGNA

Anuario estadístico de España, 1959, Madrid.

ID. — *Estadística de emisiones de capital*, año 1957, 1958. Presidencia del Gobierno, Instituto Nacional de Estadística.

ID. — *Estadística de préstamos hipotecarios correspondiente a los años 1955, 1956-57 y 1958*, Presidencia del Gobierno, Instituto Nacional de Estadística, Madrid, 1959.

ID. — *Movimiento natural de la población de España, años 1951, 1952 y 1953*, Tomo II^o, *Clasificaciones especiales de la defunciones*, Madrid, 1958.

SVEZIA

Nationalbudget 1959 with a summary in English, Stockholm, 1959.

ID. — *Statistiska Centralbyran*, 100 ar, Stockholm, 1959.

SVIZZERA

Charge fiscale en Suisse 1958, Statistiques de la Suisse, 305 fasc., Berne, 1959.

ID. — *Elections au Conseil National 1955*, Bureau fédéral de la Suisse, Contributions à la Statistique suisse, 29 fasc., Berne 1958.

ID. — *Finances et impôts de la Confédération, des Cantons et des Communes 1957-58*, Statistiques de la Suisse, 306 fasc., Berne, 1959.

ID. — *Impôt fédéral pour la défense nationale, VIII^o période*, Statistiques de la Suisse, 302 fasc.

ID. — *Mouvement de la population en Suisse : 1949-52, 1953 et 1954, 1955 1956. Tableaux*, Statistiques de la Suisse, 276, 290, 297, 303 fasc., Berne, 1957, 1958, 1959.

ID. — *IV Recensement Fédéral des Entreprises*, 25 août 1955, Vol. I. *Utilisation du sol en Suisse 1955*, Bureau fédéral des Statistiques de la Suisse, 307 fasc., Berne, 1959.

ID. — *Subventions fédérales et parts des Cantons 1957*, Berne, 1959.

U.S.A.

Annual report of the Smithsonian Institution, 1957.

ID. — « *Fertility* ». *Special Reports prepared under the supervision of H. G. Brunsman*, 1950 U.S.A. Census of Populations, 1955.

- JAFFE A. J., « *Handbook of statistical methods for Demographers* ». *Selected problems in the analysis of census data* (Preliminary edition - Second printing) U.S.A. Department of Commerce, 1951.
- ID. — Report on « Health and Demography », 44 charts reprinted from publications of Bureau of the Census and the Public Health Service, 1956.
- ID. — STATISTICAL LABORATORY OF JOWA COLLEGE, *Annual Report 1957-58*, Research Reaching Service Consultation, Agricultural Experimental Station, U.S. Bureau of Agricultural Economics.
- ASSOCIATION INTERNATIONALE DE LA SECURITE SOCIALE. — II^e *Conference Internationale des Actuaïres et Statisticiens de la Sécurité sociale*, Rome, 22-28 octobre 1959, voll. I^o et II^o.
- FOOD AND AGRICULTURE ORGANISATION OF THE UNITED NATIONS. — *Rapport de la deuxième réunion du groupe de travail ad hoc des problèmes de Sociologie rurale en Europe*, Roma, 15-17 aprile 1959.
- INSTITUT INTERNATIONAL DE STATISTIQUE. — *Bulletin de l'Institut International de Statistique*, 30^e Session, Stockholm 1957, *Actes*, Tome 36, I^o livraison, 4^e livraison, Stockholm, 1959.
- (The) NATIONAL INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH. 16th *Annual Report 1958*, London, 1959.
- ORGANISATION EUROPEENNE DE COOPERATION ECONOMIQUE. — *Bulletins Statistiques*, Commerce Extérieur.
- Série I. — *Commerce total par origine et destination*, 1958 : n. 6, décembre 1959. 1958 : n. 1, février, n. 2, avril, n. 3 juin, n. 4, août, n. 5, octobre, n. 6, décembre, 1959.
- Série II. — *Commerce par zones et par catégories de produits*, 1958 - janvier/juin, janvier/septembre, janvier/décembre, 1958.
- Série III. — *Commerce par produits*, 1957-1958. 1958 : janvier-mars/juin, janvier-juin/septembre 1958.
- Série IV. — *Commerce extérieur par produits et zones d'origine et de destination : Autriche, Belgique-Luxembourg, Grèce, Norvège, Portugal, Suisse, Turquie*, janvier-mars, janvier-juin 1958.
Canada, Danemark, Etats Unis d'Amérique, Italie. Janvier-juin 1958.
Autriche, Belgique, Canada, Danemark, Etats Unis d'Amérique, Grèce, Irlande, Islande, Luxembourg, Portugal, Royaume Uni, Suède, Suisse, Turquie. Janvier-septembre, janvier-décembre 1958.
Ensemble des pays membres 1958, Paris 1959.
- ID. — *Définitions et méthodes. II. Produits industriels*, Paris, 1959.
- ID. — *Echange par produits. Vol. I, Exports, Vol. II, Imports*, janvier-mars 1959. Série C.

- Id. — *Les engrais. Production, consommation, prix et commerce en Europe* 8^e étude 1956-1959.
- Id. — *Main d'œuvre 1900-1958*, août, 1959.
- Id. — *Résumés analytiques* janvier-juin 1959, Série B
I, Allemagne (R.F.), Grèce, Turquie.
II, Belgique-Luxembourg, France, Islande, Suède.
- Id. — *Statistiques de l'agriculture et de l'alimentation*, Paris, 1959.
- Id. — *Statistique de base de l'énergie pour les pays de l'OECE 1950-57*, 1958.
- Id. — *Statistiques générales : comptes nationaux 1959*, nn. 1, 2, 4, 5, 6, janvier, mars, juillet, septembre, novembre, 1959.
- Id. — *Statistiques de pâtes et papiers*, I^o semestre 1958, 1958.
- Id. — *Structure du commerce intraeuropéen. Les échanges de produits en 1958*.

UNITED NATIONS

- Aménagements hydroélectriques existants, en construction ou en projet sur le Danube*, Genève, 1959.
- Id. — *Annual bulletin of electric energy Statistics for Europe*, Vol. IV, june 1959, Geneva.
- Id. — *Annual bulletin of gas Statistics for Europe*, Vol. III, 1959, Geneva.
- Id. — *Annual bulletin of Transport Statistics for Europe*, tenth year, 1958, Genève, 1959.
- Id. — *(Le) commerce de produits agricoles de l'URSS et d'autres pays de l'Europe orientale de 1953 à 1957, étude statistique*, Genève, 1959.
- Id. — *Demographic yearbook 1958*, New York, 1958.
- Id. — *Electrification rurale*, Vol. IV, Genève, 1959.
- Id. — *(Le) marché européen de l'acier en 1958*, Genève, 1959.
- Id. — *Mécanisation de l'agriculture. Machines à traire et méthodes de traite*, Genève, 1958.
- Id. — *Mécanisation de l'agriculture. Le problème général des transports à la ferme*, Genève, 1959.
- Id. — *Perspectives de l'énergie nucléaire comme nouvelle source d'énergie*, Genève, 1959.
- Id. — *Rapport annuel* (25 avril 1958 - 6 mai 1959), Conseil économique et social, Procès verbaux officiels : XXVIII session, supplément n. 3.
- Id. — *Rapport sur les problèmes économiques du stockage souterrain du gaz*, Commission pour l'Europe, 1959.
- Id. — *La situation de l'énergie électrique en Europe en 1957*, Genève, 1959.
- Id. — *Statistiques des accidents de la circulation routière en Europe en 1957*. 1959.

UNIVERSITÀ

Università degli Studi di Roma — Facoltà di Scienze Stat. Dem. ed Att.,
Atti del Seminario di Sociologia, anno accademico 1955-56, Sezione di
Sociologia, nuova serie, Vol. I.

University College of Wales, Aberystwyth. Welsh Plant Breeding
Station, *Report* (1^o October 1956 to 30th September 1958), Plas Gogend-
dan, Aberystwyth, 1959.

Libri, Annuari, Opuscoli,
Estratti, ecc.

Books, Yearbooks,
Pamphlets, Reprints etc.

Libros, Anuarios, Folletos, Extractos ecc.

Livres : Annaire, Brochures
Extraits, etc.

Bücher, Jahrbücher,
Broschuren, usw.

AGRICOLTURA

NATIONS UNIES. — *Electrification rurale*. Préparé par le Secrétariat de la Commission économique pour l'Europe, Genève, 1959.

JANNACCONE AMEDEO. — *Sulla modificazione del rapporto numerico tra i sessi nella Canapa*. (Ricerche Sperimentali). Estratto dagli « Annali della Facoltà di Agraria della Università di Napoli. Serie III, Vol. X, 1939.

MANGELSDORF PAUL C. — *Wheat*. By the Review « Scientific American » Genealogy of Wheat. July 1953.

ID. — *The Mutagenic Effect of Hybridizing Maize and Teosinte*. Reprinted from Cold Spring Harbor Symposia on Quantitative Biology. Volume XXIII, 1958.

ID. — *Reconstructing the Ancestor of Corn*. Reprinted from Proceedings of the American Philosophical Society. Vol. 102, No. 5, October, 1958.

NATIONS UNIES ; COMMISSION ECONOMIQUE POUR L'EUROPE. — *Mécanisation de l'Agriculture. Le problème général des transports à la ferme*. Genève.

ID. — *Mécanisation de l'agriculture. Machines à traire et méthodes de traite*. Genève.

PETERSON PETER A. — *The Effect of temperature on the Mutation Rate of a Mutable Locus in Maize*. Reprinted without change of paging from the Journal of Heredity Washington, D.C. Vol. XLIX, No. 3, May-June 1958.

ZANCHERI RENATO. — *Prime ricerche sulla distribuzione della proprietà fondiaria nella pianura bolognese (1789-1835)*.

ALIMENTAZIONE

ORGANISATION DES NATIONS UNIES POUR L'ALIMENTATION ET L'AGRICULTURE. — *Rapport de la Réunion du groupe de travail ad hoc des problèmes de sociologie rurale en Europe*. Avril 1958, Rome - Italie.

ID. — *Rapport de la deuxième réunion du groupe de travail ad hoc des problèmes de sociologie rurale en Europe*. Avril 1959, Rome - Italie.

ANTROPOLOGIA

- McILWRAITH T. F. — *Bibliography of Canadian Anthropology for 1955*. (From Bulletin no. 147, Annual Report of the National Museum for the Fiscal Year 1955-56). Canada.
- LIVINGSTONE FRANK B. — *Anthropological Implications of Sickle Cell Gene Distribution in West Africa*. In «American Anthropologist», Vol. 60, No. 3, June 1958.
- MOOR-JANKOWSKI JAN K. and HUSER H. J. — *Sero-Anthropological Investigations in the Walser and Romansh Isolates of the Swiss Alps*. In «International Society of Blood Transfusion». Sept. 3-5 1956.
- POLLITZER W. S. — *The negroes of charleston (S.C.); a study of hemoglobin types, serology and morphology*. In «The American Journal of Physical Anthropology». Vol. 16, No. 2, June 1958.
- SCHLAGANHAUFEN (VON) OTTO. — *Mitteilungen uber einen Riesen*. Separatabdruck aus: Bulletin der Schweizerischen Gesellschaft fur Anthropologie und Ethnologie. 35. Jahrgang, 1958/59, S. 23-37, Buchdruckerei Buchler & Co., Bern, 1959.
- SHAPIRO HARRY L. — *A reprint from American anthropologist*. The History and development of physical anthropology. Vol. 61, No. 3, June 1959.
- SPUHLER J. N. — *Formation et Transformation des Races : Origine de l'Espèce humaine Nouvelles Théories*. Guy Dingemans. (Am. Anthro. 60, april 1958).

ARCHEOLOGIA

- CRUXENT J. M. and ROUSE IRVING. — *An archeological chronology of Venezuela*. Social Science Monographs, VI. Vol. I, Washington, 1958.

ASSICURAZIONI

- COPPINI MARIO ALBERTO. — *La piccola riforma dell'I.N.A.M.*. Estratto dalla Rivista: «I Problemi della Sicurezza Sociale». Anno XIV, n. 5, settembre-ottobre 1959. I.N.A.M., Roma.
- ID. — *Una importante esperienza italiana sulla eliminazione delle rendite derivanti da infortunio*. Estratto: dal «Giornale dell'Istituto Italiano degli Attuari». Anno XXI, 1958, Istituto Italiano degli Attuari, Roma, 1958.
- ID. — *Che cosa ci separa dalla sicurezza sociale*. Estratto: dalla Rivista: «I Problemi della Sicurezza Sociale». Anno XIV, n. 2, marzo-aprile 1959, I.N.A.M., Roma.
- ID. — *Per una Progettazione razionale dei Fondi Pensioni*. Estratto dalla Rivista: «Previdenza Sociale». Gennaio-febbraio 1959, I.N.P.S., Roma.
- ID. — *De l'élimination des rentes dérivant d'accidents du travail*. Année 1958-1959.
- ISTITUTO NAZIONALE DELLA PREVIDENZA SOCIALE. — *Sintesi dei principali sistemi previdenziali del mondo*. Servizio affari generali - Ufficio Studi e rapporti culturali, I.N.P.S., Roma, 1959.
- ISTITUTO NAZIONALE DELLA PREVIDENZA SOCIALE. (Servizio Statistico Attuariale). — *Statistiche delle Pensioni dell'Assicurazione obbligatoria, Invalidità, Vecchiaia e Superstiti* al 31 Dicembre 1958. Arti grafiche Pannetto & Petrelli, Spoleto, Roma 1959.

JECKLIN H. — *Simplification de la méthode pour le calcul des réserves mathématiques en assurance-vie* in : « Boletim do instituto do actuários Portugueses », N. 12, 1956.

BIBLIOGRAFIA.

GEDDA LUIGI. — *Due Pontefici : Pio XII - Giovanni XXIII*. A cura del Banco di Roma. Roma, 1959.

BIOGRAFIA

AMOROSO LUIGI : *Einaudi scienziato ed artista*. IPSI, Pompei 1959.

FORTUNATI P., GIANNONE A. — *Nel ricordo di Mario Saibante*. In « Statistica », Ott.-dic. 1958, Bologna 1958.

PELLA GIUSEPPE, J. D. ZELLERBACH. — *John Foster Dulles*. (Discorsi commemorativi tenuti a Roma, il 7 luglio 1959 nella sede del Banco di Roma. A cura del Banco di.

BIOLOGIA

BRYAN JOHN H. D., GOWEN JOHN W. — *The Effects of 2560 r OF X-Rays on Spermatogenesis in the Mouse*. Reprinted from Biological Bulletin Vol. 114, No. 3, pp. 271-283, June 1958, Printed in U.S.A.

LANE PRISCILLA W., DICKIE MARGARET M. — *The Effect of Restricted food intake on the life span of genetically obese mice*. Reprinted from « The Journal of Nutrition ». Vol. 64, No. 4, April 1958 Copyright. 1958 - The Wistar Institute of Anatomy and Biology.

KJER POUL. — *Infantile optic atrophy with dominant mode of inheritance*. Opera ex domo biologiae hereditariae humanae. Universitatis hafniensis Vol. 42. Ejnar Munksgaard Copenhagen, 1959.

MILBANK MEMORIAL FUND. *Thirty years of research in Human fertility : retrospect and prospect*. Milbank Memorial Fund, New York, 1959.

WEATHRWAX PAUL. — *Indian Corn in old America*. Reprinted from « The Quarterly Review of Biology. » Volume 30, No. 1 ; March 1955. Printed in U.S.A.

CALCOLO DELLE PROBABILITÀ

BIRNBAUM Z. W. and MCCARTY R. C. — *A distribution-free upper confidence bound for $Pr(Y < X)$, based on independent samples of X Y* . In « The Annals of Mathematical Statistics ». Vol. 29, No. 2, June 1958.

DAVID HERBERT T. and KRUSKAL WILLIAM H. — *The wagr sequential t-Test reaches a decision with probability one*. In « The Annals of Mathematical Statistics ». Vol. 27, No. 3, sept. 1956.

FELLER WILLIAM. — *An Introduction to Probability Teory and its Applications*. Vol. I, John Wiley & Sons, Inc. New York, 1957.

DEMOGRAFIA

- ALBERTI SALVATORE. — *Condizioni di abitazione e stati morbosi*. Inchiesta schermografica su famiglie assegnatarie di alloggi INA-Casa. Ente Gestione Servizio Sociale, Roma, Gennaio 1958.
- BARCLAY GEORGE W. — *Techniques of population analysis*. John Wiley & Sons, Inc. Publ. New York, 1958.
- CAPPIERI MARIO. — *Lineamenti di demografia di alcuni gruppi razziali isolati*. Università degli Studi di Roma, Fac. Scienze Statistiche Demografiche ed Attuariali. Ed. I, 1952. Libreria Eredi Virgilio Veschi, Roma.
- Fifth public opinion survey on birth control in Japan*. Population Problems Series No. 16. Tokyo, Japan, 1959.
- Health and Demography*. Report on « Health and Demography ». 44 charts reprinted from publications of Bureau of the Census and the Public Health Service U.S. Department of Health, Educations and Welfare. October 1956.
- IMPRESCIA CARMELO. — *Il controllo delle nascite e la natalità*.
- GUTMAN ROBERT. — *Birth and Death Registration in Massachusetts 1639-1900*. Reprinted from The Milbank Memorial Fund Quarterly. Vol. XXXVI, No. 1 and 4; and Vol. XXXVII, No. 3 and 4 1959. Milbank Memorial Fund. New York, 1959.
- ISTITUTO DI STATISTICA. UNIVERSITÀ DI BOLOGNA. — *Evoluzione Demografica ed insediamento umano nell'Emilia-Romagna*. Bologna. Tipografia Compositori, 1959.
- McILWRAITH T. E. — *The relation of man to nature through the ages*. In « The Canadian Field-Naturalist » Vol. 69, July-Sept. 1956, No. 3, Ottawa, 1955.
- KUCZYNSKI R. R. — *A demographic survey of the British Colonial empire*. Vol. III, Oxford University Press, London, 1953.
- MARBACH GIORGIO. *La concentrazione del carattere : intervallo tra matrimonio e nascita del primogenito*. In « Rivista di Politica Economica ». Roma, Nov. 1958.
- MAROTTA MICHELE. — *Rassegna demografica della Sardegna*. Estratto da Ichnusa 1956, III, 11 ; 12, Gallizzi, Sassari 1956.
- MORTARA GIORGIO. — *La mortalità nel sud e nel nord dell'Italia*. Roma, settembre 1958.
- ID. — *Tavole di sopravvivenza per il sud e il nord dell'Italia*. Roma, Dicembre 1958.
- ID. — *Previsione demografica e previsione economica*. In « L'Industria », No. 3, 1958. Casa Editrice l'Industria, Milano, 1958.
- NADDEO ALIGHIERO. — *Su alcune caratteristiche demografiche dell'isola di Ischia*. In « Minerva Medica ». Vol. 50, 1959. Edizioni Minerva Medica.
- SIGMOND RAIMONDO. — *Il Problema Demografico*. — Morcelliana, Brescia 1958
- UNITED NATIONS. — *Department of Economic and Social Affairs*. « Case Studies of Arrangements for evaluation and utilization of Population census result ». Report I, The Republic of the Sudan United Nations - New York, 1959.
- WHELPTON P. K. and KISER CLYDE V. — *Social and psychological factors affecting fertility*. Vol. Five. Milbank Memorial Fund, New York, 1958.

DIRITTO

- CANCELLI FILIPPO. — *Saggio sul concetto di « officium » in diritto romano*. Dott. A. Giuffrè Editore, Milano, 1958.
- MONTENEGRO BACA JOSE. — *Jornada de Trabajo y Descansos Remunerados*. Estudio Doctrinario y Critico de las Disposiciones Legales y Jurisprudenciales sobre la Materia. Tomo I, Grafica Jacobs S. A. Lima, 1959.
- ID. — *Jornada de Trabajo y Descansos Remunerados*. Legislacion - Jurisprudencia - Indices. Tomo II, Tip. Salas é Hijos S. A., Lima, 1959.
- NIRCHIO GIUSEPPINA. — *Sociologia giuridica e scienza del diritto* (A proposito di un recente giudizio del Carnelutti). In « *Il Circolo Giuridico* », Anno 1959, Tip. Michele Montaina, Palermo, 1959.
- PEPIN EUGENE (Centro per lo Sviluppo dei trasporti aerei). — *L'Evolution de l'enseignement du Droit aérien dans le monde*. Roma, 1959.
- VINCENZI G. — FLAMMIAR. — *Le Norme Transitorie per garantire minimi di trattamento economico e normativo ai lavoratori*. Estratto dalla Rivista: « *Massimario di Giurisprudenza del Lavoro* ». IV Serie, Anno XXXII, N. 5-6, 1959, Tipogr. Delle Terme. Roma.

ECONOMETRICA

- FOSSATI ERNALDO. — *Il significato della ricerca econometrica*. Genova, 11 dicembre 1958.
- MALINSCHI V. — *L'analyse statistique du niveau d'existence dans les études monographiques*.
- NATAF ANDRÉ. — *Forme Réduite d'Agrégats de Consommation dans le Cadre de la Théorie des Choix* No. 4. Extrait des: « *Publications de l'Institut de Statistique de Paris* ». Vol. VII, Fasc. 1, 1958.
- WOLD HERMAN O. A. — *Ends and means in econometric model building*. For limited circulation. First draft, February, 1959.

ECONOMIA

- ALLAIS MAURICE. — *Fondements théoriques, Perspectives et Conditions d'un Marché Commun Effectif*. N. 2. In « *Le Marché Commun et ses Problèmes* », Jan. Fév., 1958, Editions Sirey, Paris, 1958.
- ALLEGRI ALBERTO. — *Il commercio e i Comuni d'Italia*. « *Memoriale dei Comuni* », Mag. 1959, Empoli, 1959.
- AKERMAN JOHAN. — *L'aspect structurel*. In: « *Revue Economique* », N. 6, Nov. 1954, Lib. Armand Colin, Paris, 1954.
- ID. — *Structural limits in economic development*. Overdruk Economist, Jaargang, 1949.
- ID. — *L'analyse structurelle des variations économiques*. Extrait du « *Bulletin de l'Institut de Recherches Economiques et Sociales de l'Université de Louvain XIV, Année, N° 3-5 déc. 1948*. Institut de Recherches Economiques et Sociales Louvain, 1948.
- AMOROSO LUIGI. — *L'Orizzonte Economico nel pensiero di Maffeo Pantaleoni*. Lezione tenuta all'Università di Bari il 22 aprile 1959. Annali dell'Istituto di Statistica dell'Università di Bari, 1959.

- ASSOCIAZIONE FRA LE SOCIETÀ ITALIANE PER AZIONI. — *La disciplina fiscale degli ammortamenti nei principali stati esteri*. Quaderni XXIV, Roma 1959.
- BANCO DI SICILIA. — *Lineamenti e prospettive di sviluppo dell'economia siciliana*. Banco di Sicilia, Palermo.
- BARNA T. — *Investment in industry has Britain lagged?* National Institute of economic & social research London Reprint Series No. 10. In «The Banker», April 1957.
- BERGSON ABRAM and HEYMANN HANS, Jr. — *Soviet National Income and Product 1940-48*. Columbia University Press, New York, 1954.
- BJERVE PETER JAKOB. — *Interpretation of deviations between national budget and national accounts for norway*. Almqvist & Wiksells, Uppsala, 1957.
- BOZZOLA GIAN BATTISTA. — *I Problemi non risolti del nostro sviluppo industriale*. Estratto da: «La rivoluzione industriale nelle regioni meridionali conseguenze umane e sociali «Fattore Umano», N. 5, maggio 1957.
- BRESCIANI-TURRONI C. — *Corso di economia politica*. Problemi di politica economica. Vol. II Editore A. Giuffrè, Milano, 1957.
- ID. — *Corso di economia politica*. Teoria generale dei fatti economici. Vol I, A. Giuffrè, Editore, Milano, 1956.
- BROWN J. A. C. — *Seasonality and elasticity of the demand for food in Great Britain since derationing*. Reprint Series, No. 148 in «Journal of Agricultural Economics», Dec. 1958.
- CAIRNCROSS ALEC. — *The international bank for reconstruction and development* No. 33, March. 1959, Princeton University, Princeton 1959.
- COLE W. A. — *The measurement of industrial growth*. N. 151. The Economic History Review. S. II, Vol. XI, No 1, 1958.
- D'IPPOLITO TEODORO. — *I costi di produzione e di distribuzione*. Principi e procedimenti di determinazione. Abbaco editore, Roma, 1958.
- ID. — *I costi di produzione e di distribuzione*. Principi e procedimenti di determinazione. Tomi I, Testo, Abbaco Editore, Roma, 1958.
- D'IPPOLITO GIOACCHINO. *Economia politica*. Corso introduttivo (Cinematica economica) 1953-1954. Università degli Studi di Roma Fac. di scienze statistiche Dem. ed Attuariali. Libreria Eredi Virgilio Veschi, Roma, 1958.
- DOMAR D. EVSEY. — *Essays in the Theory of Economic Growth*. Oxford University Press, Oxford, 1957.
- FERAUD LUCIEN, SOLARI LUIGI. — *Sur les Attentes à l'entrée de différents services*. Recherche Opérationnelle. Extrait du Volume XIV des publications de la Faculté des sciences économiques et sociales de l'Université de Genève. Librairie de l'Université. Georg & Cie. S. A. Genève, 1959.
- FERRO OTTONE. — *Dinamica Economica in due ambienti a differente indirizzo produttivo*. Estratto da: «Rivista di Politica Agraria». Dicembre 1959. Edizioni Agricole, Bologna.
- FORTUNATI PAOLO. — *Errori, equivoci e schemi corporativi nella regolamentazione dei mercati*. Tip. del Senato, G. Bardi, Roma, 1959.
- ID. — *Problemi di metodo e di conoscenza nella analisi politico-economica*. Tip. del Senato, Roma 1959.
- ID. — *Condizioni oggettive e impegni costituzionali di una politica economica di struttura*. Tip. del Senato, Roma, 1959.
- ID. — *Economia e politica*. Tip. del Senato, Roma 1958.
- GIANNONE A. — *Le ricerche sul reddito nazionale in Italia*. In «Rassegna Economica», Ott.-dic. 1958.

- HAYAKAWA MIYOJI. — *Distribution of Income in Japan 1905-1956*. Reprinted from « Waseda Economic Papers, No. 4, 1959. Graduate Division of Economics, Waseda University, Tokyo - Japan.
- HARTLEY H. O. — *Analytic studies of survey data*. Estratto dal Volume in onore di Corrado Gini. Tip. Fausto Failli, Roma.
- INDUSTRIALIZACION (I.A.) Y EL AHORRO DE DIVISAS EN LA ARGENTINA. — *FIAT - Delegacion Para America Latina*. Oficina De Estudios para la colaboracion economica internacional. Buenos Aires, 1959.
- ISTITUTO ITALIANO DI PUBBLICISMO. — *Saggi e Studi di Pubblicità* — Nona e Decima Serie, 1957-1958 Istituto Italiano di Pubblicità, Roma, 1959.
- JANNACCONE PASQUALE. — *Manuale di economia politica*. UTET, Torino, 1959.
- HAYES SAMUEL P. Jr. — *Comment mesurer les résultats des projets de développement*. Manuel à l'usage des missions sur le terrain. Rédigé pour l'Unesco, 1959.
- KLAIBER MANFRED. — *Si può parlare di un miracolo nell'economia germanica?* A cura del Banco di Roma, Roma, 1959.
- LOYO GILBERTO. — *Un Ensayo Acerca de la Planificacion Economica de la Republica Mexicana*. Prologo del libro « Diagnostico Economico Regional » Mexico, 1959.
- MALVESTITI PIERO. — *Discorso sulla zona di libero scambio*. A cura del Banco di Roma, Roma, 1959.
- MEDICI GIUSEPPE. — *Replia agli interventi sulla esposizione economica e finanziaria*. (Senato 31 ottobre 1958). Istituto Poligrafico dello Stato, Roma, 1958.
- MENTASTI FLAMINIO. — *Ricerche in tema di correlati empirici a periodicità inferiore all'anno per il flusso produttivo della agricoltura italiana*. Istituto Nazionale per lo studio della congiuntura. ISCO, Roma, gennaio 1959.
- MYRDAL GUNNAR. — *An International Economy Problems and Prospects*. Harper & Brothers Publishers, New York, 1956.
- NATIONS UNIES. — *Etude sur la situation économique de l'Europe en 1958*. Genève, 1959.
- ID. — *Le commerce de produits agricoles de l'URSS et d'autres pays de l'Europe Orientale de 1953 à 1957*. Etude statistique, Genève, 1959.
- ID. — *Les incidences économiques de l'automatisation en Europe*. Genève, 1959.
- PAIGE DEBORAH and BOMBACH GOTTFRIED. — *A comparison of National output and productivity of the United Kingdom and the United States*. OEEC. Paris, 1959.
- PANIZZON GAETANO. — *Elasticità della domanda di alcuni generi alimentari secondo l'esperienza Italiana*. Estratto dalla « Rivista Internaz. di Scienze Economiche e Commerciali. Anno VI (1959), n. 12, CEDAM, Padova, 1959.
- PHILLIPS A. W. — *The relation between unemployment and the rate of change of money wage rates in the United Kingdom, 1861-1957*. No. 87. In *Economica*, Vol. XXV, No. 100, 1958.
- REDDAWAY W. B. — *Some limitations of the capital-output ratio*. Reprint Series, No. 150. In « The Pakistan Economic Journal ». Vol. VIII, 1958.
- ROMANIUK KAZIMIERZ. — *Polish interindustry relations tables for 1956*.
- Situation (La) et les perspectives du marché charbonnier européen en 1958/1959*. Nations Unies, Genève, 1959.

- SOLARI LUIGI. — Vedi: FERAUD LUCIEN: *Sur les Attentes à l'entrée de différents services*. Recherche Operationelle. Genève, 1959.
- STERIOTIS P. I. — *La méthode géométrique sur l'arbitrage dans le change indirect*. Athènes, 1958.
- SAVORGNAN FRANCO. — *Nel cinquantenario della morte di Ludwig Gumpowicz*. In « Rivista Italiana di Economia, Demografia e Statistica », Vol. XIII, 1959, Soc. Abete, Roma, 1959.
- STEWART I. G. — *Input-output table for the united Kingdom*. No. 147. In « The London and Cambridge Economic Bulletin ».
- STUDENSKI PAUL. — *The Income of Nations. Theory, measurement, and analysis: past and present*. New York, University Press, 1958.
- TAGLIACARNE GUGLIELMO. — *Calcolo del reddito prodotto dal settore privato e dalla pubblica amministrazione nelle provincie e regioni d'Italia nel 1957 e indici di alcuni consumi e del risparmio bancario e postale*. In « Moneta e Credito », N. 44, Roma, 1958.
- THOMAS BRINLEY. — *The Economics of international migration*. MacMillan & Co. Ltd., London, 1958.

ECONOMIA AGRARIA

- BANDINI MARIO. — *La politica agraria italiana*. Convegno di studi « L'agricoltura e il mercato comune europeo ». Roma, 23-25 gennaio 1958 Confederazione Nazionale Coltivatori Diretti.

ECONOMIA AZIENDALE

- D'IPPOLITO TEODORO. — *L'azienda*. Prime nozioni di economia aziendale e di ragioneria. IV* edizione, Abbaco editore, Roma, 1958.
- ID. — *La contabilità in partita doppia a sistema unico e duplice ed il bilancio d'esercizio*. IV* edizione, Abbaco Editore, 1958, Roma.
- LENTI LIBERO. — *Metodi vecchi e nuovi per la gestione delle imprese*. — Articolo tratto dal quotidiano « Corriere della Sera », Martedì 20 Ottobre 1959.

ECONOMIA : BANCHE

- TAGLIACARNE GUGLIELMO. — *Calcolo del reddito prodotto dal settore privato e dalla Pubblica Amministrazione nelle provincie e regioni d'Italia dal 1952 al 1958 e Indici di alcuni consumi e del risparmio bancario e postale*. Estratto da: « Moneta e Credito », N. 48. Rivista trimestrale della Banca Nazionale del Lavoro. Roma, 1959.

ECONOMIA : FINANZA

- KRIZ MIROSLAV A. — *Gold in world monetary affairs today*. Essay in international finance No. 34, June 1959. Princeton, 1959.
- TAMBRONI FERNANDO. — *Relazione generale sulla situazione economica del paese* (1958). Roma, 1959.
- ID. — *Il bilancio economico nazionale per il 1958*. Roma, 1959.
- ID. — *Esposizione economica e finanziaria per una politica di sviluppo*. Roma, 1959.

ECONOMIA : INDUSTRIA

Die höhere schule in der technisch industriellen arbeitswelt. Vorträge und Diskussionen zur 12. Jahreshauptversammlung. Goslar, Am 18. und 19. Juni 1959.

PEDRONI F. — *Il Concetto integrale della produttività e le concrete possibilità di misurazione.* L'Eco della Stampa, Milano, Agosto 1959.

SCOTTO ALDO. — *Problemi fiscali del mercato comune nei confronti di una associazione tra i paesi dell'O.E.C.E.* Relazione presentata al IX Convegno di Studi di Economia e Politica Industriale. Sanremo, 15-16-17 nov. in « Rivista di Politica Economica ». Anno XLIX, III S., Fasc. I - 11 gen.- feb. 1959.

ECONOMIA : MONETA

PAPADAKIS J. — *Inflation, employment economic development and free economy.* Buenos Aires : 1959.

ECONOMIA : PREZZI

ALLAIS MAURICE. — *Le Système des Prix et la Concurrence dans le Marché Commun de la C.E.C.A.* N° 1. In « Actes Officiels du Congrès d'Etudes sur la C.E.C.A. ». Milan, 1957, C.N.R.S. 1957.

ECONOMIA : STORIA

BEER SERGIO. — *La divulgazione della scienza nella stampa periodica a grande diffusione.* U.S.P.I., Roma, 30-31 marzo 1957.

ETNOLOGIA

BOCCASSINO RENATO. — *Etnologia religiosa.* — Storia e scienza delle religioni. Società Editrice Internazionale, Torino, 1958.

TUCCI G. — *Note sur l'orientation actuelle des Etudes Ethnographiques en Italie.* Dal : N. 93, Octobre 1959 (Serie M, N. 5)) di : Recherches et Dialogues Philosophiques et Economiques. I.S.E.A., Paris.

EUGENICA

HERNDON NASH C. — *Procedures for referral to heredity counselors.* In : « Eugenics Quarterly ». Vol. 5, No. 1, March 1958.

MILBANK MEMORIAL FUND. — *Thirty Years of Research in Human Fertility.* Retrospect and Prospect. Papers presented at the 1958 Annual Conference of the Milbank Memorial Fund, Held October 22-23 1958, at the New York Academy of Medicine, Part II. Milbank Memorial Fund, New York, 1959.

SHAPIRO HARRY L. — *Eugenics and Future Society.* Reprinted from Eugenics Quarterly ». Vol. 6, No. 1, March 1959, Published by American Eugenics Society. N.Y.C.

FILOSOFIA

BELLO ANDRES. —

Vol. I. *Poesias*.

Vol. III. *Filosofia*.

Vol. IV. *Gramatica*.

Vol. V. *Estudios Gramaticales*.

Ed. Ministerio de Educacion, Caracas, 1951.

FISIOLOGIA

BOUTOURLINE YOUNG H., MILANI COMPARETTI A., MORI FERRARA G., NICOLETTI I. e ZOLI A. — *Studi sulla fisiologia del bambino a scuola*. In: « Rivista di clinica pediatrica ». Vallecchi Editore, Firenze.

IGIENE E SANITÀ

BOUTOURLINE YOUNG H. — *European Research in child psychology*. Vol. 5, No. 3.

ID. — *Su alcuni aspetti della salute fisica e mentale di un gruppo di bambini di scuola elementare e media di Firenze*. In « Annali della Sanità Pubblica ». Vol. XVI, Fasc. II, Marzo-aprile 1955. Tip. Regionale, Roma 1955.

BOUTOURLINE YOUNG H., NICOLETTI IVAN. — *Un'analisi delle deviazioni della salute trovate in bambini delle Scuole Medie di Firenze*. In « Minerva medica », Anno XLVII, Vol. II, N. 55 (11 Luglio, 1956), Edizioni Minerva Medica, Torino.

GENETICA

ALLEN GORDON and FIRSCHEIN LESTER I. — *The Mathematical Relations Among Plural Births*. In « American Journal of Human Genetics », Vol. 9, Dept. 1957.

ALSTROM CARL HENRY. — *First-Cousin Marriages in Sweden 1750-1844 and a Study of the Population Movement in Some Swedish Subpopulations from the Genetic-Statistical Viewpoint*. Laboraty No. 2 for: « Human Genetics » The Psychiatric Clinic of the Caroline Institute, Stockholm - Sweden. Acta Genetica et Statistica Medica. Vol. 8, N. 3-4 1958, Basel S. Karger, New York, 1958.

ASCNER B. M. and GARTLER S. M. — *Physiologic variation of renal function in twins: diuresis after water intake*. In « Acta Geneticae Medicae et Gemellologiae », Vol. VII, Januari, 1958.

CUCCO ALFREDO. — *La donna e la maternità*. Gli oratori del giorno. Roma, maggio 1959.

EDWARDS A. W.F. — *An analysis of geissler's data on the human sex ratio*. In Reprinted from: « Annals of Human Genetics, Vol. 22, Part. I.

HOWARD, G. BRUNSMAN (Special Reports prepared under the supervision of) — *Fertility*. 1950, United States Census of Population. United States Government Printing Office, 1955.

- FREIRE-MAIA N. — *Consanguineous Marriages in Brazil: 1) Structure of Such Marriages. 2) Factors of Geographic Distribution*. Reprinted from «Eugenics Quarterly». Vol. 5, No. 2, June 1958. Published by American Eugenics Society, N.Y.C.
- ID. — *Inbreeding in Brazil*. Reprinted from: «American Journal of Human Genetics». Vol. 9, No. 4, December 1957. Printed in U.S.A.
- FREIRE-MAIA N., QUELCE SALGADO A., AMUNDSEN KOEHLER R. — *Hereditary Bone Aplasias and Hipoplasias of the Upper Extremities*. — Acta Genetica 9: 33-40 (1959). Vol. 9, No. 1 (1959) Separatum. Printed in Switzerland.
- FREIRE-MAIA N., QUELCE SALGADO A., FREIRE-MAIA A. — *Hand Clasping in different Ethnic Groups*. Reprinted from: «Human Biology». December, 1958, Vol. 30, No. 4.
- FREIRE-MAIA N. — *Inbreeding Levels in different countries*. Reprinted from «Eugenical News», Vol. IV, No. 3, September, 1957. Published by American Eugenics Soc. N.Y.C.
- ID. — *Effect of inbreeding levels of populations on incidence of hereditary traits due to induced recessive mutations*. Reprinted from: «Effect of Radiation on Human Heredity». World Health Organization, Geneva, 1957.
- FUNG-SUI-TONG, STADLER IANICE, GOWEN JOHN W. — *Sex Differentiation as Affected by Major Genes and Gene Dosages*. Genetics Department, Iowa State College, Ames, Iowa, U.S.A. Estratto: «Journal Paper» No. J. 3425 of the «Iowa Agricultural and Home Economics Experiment Station, Ames, Iowa. Project No. 1187. Printed in Canada.
- GOWEN JOHN W., SUI-TONG CHAN FUNG. — *Determination of Sex through Genes in a Major Sex Locus in Drosophila Melanogaster*. Reprinted from «Heredity». Vol. 11, Part 3, pp. 397-402. December 1957.
- GOWEN JOHN W. — *Genetic consideration of Leukosis tumor Growth in Poultry*. Reprinted from Annals of New York Academy of Sciences. Vol. 71, Art. 6, September 30, 1958.
- GOWEN J. W., STADLER J. — *The Totipotency of All Cells for Typhoid Resistance Salmonella typhimurium, in Mice*. Reprinted from: Supplement Volume of «Cytologia» International Journal of Cytology. Proceedings of the International Genetics Symposia, 1956: 410-415. Issued July 30, 1957. Printed in Japan.
- GRAF G. E. — *Chromosome Breakage Induced by X - Rays, Maleic hydrazide and its derivatives*. Reprinted without change of paging from the journal of Heredity Washington, D.C. Vol. XLVIII, No. 4, July-August, 1957. Copyright 1957 by the American Genetic Association.
- HALLGREN BERTIL. — *Retinitis Pigmentosa combined with Congenital deafness; with vestibulo-cerebellar ataxia and mental abnormality in a proportion of cases. A Clinical and Genetico-Statistical Study*. Acta Psychiatrica et neurologica Scandinavica. Supplementum 138, Vol. 34, 1959. Ejnar Munksgaard. Copenhagen, 1959.
- NEEL JAMES V. — *The Detection of the Genetic Carriers of Hereditary Disease* In: «American Journal of Human Genetics». Vol. I, No. 1 Sept. 1949.
- OSBORNE RICHARD H. and DE GEORGE FRANCES V. — *Selective Survival in Dizygotic twins in Relation to the ABO Blood Groups*. In: «American Journal of Human Genetics. Vol. 9, No. 4, Dec. 1957.
- OSBORNE RICHARD H. — *Hereditary and environmental factors in body build: a study of 100 pairs of adult twins in*: «Dissertation Abstracts». Vol. XVI, No. 7, 1956.

- ROSIN S., MOOR-JANKOWSKI J. K. und SCHNEEBERGER MARIA. — *Die derivität in bluterstamm von tenna (hämphilie B)* in : « Acta genetica et statistica medica ». Vol. 8, No. 1, 1959.
- SADANAGA K. — *Cytological Studies of Hybrids involving Triticum durum and Secale cereale*. I Alien addition races in tetraploid wheat. Estratto : « Journal Paper No. J-3154 of the Iowa Agricultural Experiment Station ». Ames, Iowa. Project No. 1181.
- SCHWARTZ DREW. — *A new temperature-sensitive allele at the sticky locus in maize* in : « Journal of Heredity ». Vol. XLIX, No. 4, July-August 1958.
- SCHWARTZ D. and MURRAY C. B. — *A Cytological Study of Breakage-fusionbridge Cycles in Maize Endosperm* in : « Cytologia » Proceedings of the International. Genetics Symposia, 1956.
- SCHWARTZ DREW. — *Deoxyribonucleic Acid Side-Chain Model of the Chromosomes* in : « Nature ». Vol. 181, April 19-58.
- ID. — *On The Stabilization of a Ring Chromosome in Maize* in : « Genetics ». Vol. 43, No. 1, January 1958.
- SCHULL WILLIAM J. — *A Note Consanguineous Marriages in the Cities of Hiroshima and Nagasaki* in : « The Japanese of Human Genetics ». Vol. 3, No. 2, 1958.
- SCHULL WILLIAM J. and NEEL JAMES V. — *Radiation and the Sex Ratio in Man* in : « Science ». August 15-1958, Vol. 128, No. 3320.
- SCHULL WILLIAM J. — *Empirical Risks in Consanguineous Marriages : Sex Ratio, Malformation, and Viability* in : « American Journal of Human Genetics ». Vol. 10. No. 3, Sept. 1958.
- ID. — *Consanguinity and the etiology of congenital malformations* in : « Pediatrics ». Vol. 23, No. 1, Part II, Jan. 1959.
- TESAURO G. (edited by). — *Proceedings of the Second World Congress on Fertility and Sterility. Vol. I and II*. International Fertility Association. Institute of Clinical Obstetrics and Gynecology. Tipografia poliglotta Vaticana, 1958.

GEOGRAFIA

- BOLTE HENRY EDWARD. — *Australia : terra di progresso e di prosperità*. Discorso pronunciato a Roma il 15 Luglio 1959, nella Sede del Banco di Roma. A cura del Banco di Roma. 1959.
- ESTADOS UNIDOS MEXICANOS. — *Secretaria de Industria y Comercio Direccion General de Estadistica*. « Mexico en Cifras ». Mexico, 1959.
- ID. — *Secretaria de Industria y Comercio*. « Censos Agropecuarios » : 1) Totales Comparativos en 1930-1940-1950. 2) Por Entidades y Distritos Economico Agricolas en 1950. Mexico, 1959.
- GAJDUSEK CARLETON D. — *Geographical Background to the Serological Survey of the Middle East*.
- INFORMATION DEPARTMENT OF THE MINISTRY OF TRASPORT. — *History of the Delta Plan*. April 1958.
- MARQUEZ MIRANDA FERNANDO. — *Región Meridional de América de Sur*. Periodo Indígena. Instituto Panamericano de Geografía e Historia, Pub. num. 178, Mexico, 1954.
- PONENCIA (Centro de Investigaciones Agrarias). — *Modificaciones necesarias a la Legislación Agraria y Formas de Organización Campesina*. Mexico, 1959.

SCHLAGINHAUFEN (VON) OTTO. — « *Menschentypen und Rinderrassen in ihrer geographischen Verteilung in der Schweiz.* ». Ausgegeben am 27 Juni 1959. Buchdruckerei Gebr. Fretz AG, Zurich, 1959.

Un'opera da titani. Lo sbarramento e il prosciugamento dello « Zuiderzee ». Tip. C. Huig Zaandam, Olanda.

VERBAEK C. L. — *Topographical and Archaeological Investigations in the Norse Settlements in Greenland.* A Survey of the Work of the last 10 years. Reprint from: Thridji Vikingafundur Third Viking Congress. 1958.

GEOGRAFIA ECONOMICA

NATIONS UNIES. — *Amenagements hydro-électriques existants, en construction ou en projet sur le Danube.* Genève.

LETTERATURA

BELLO ANDRES. — *Estudios Filologicos.* Principios de la ortologia y métrica da la Lengua castellana y otros escritos. Vol. VI-X-XII-XIII, Ediciones del Ministerio de Educacion, Caracas - Venezuela, 1954, 1955.

ID. — *Temas de critica literaria.* Vol. IX, Ministerio de Educacion, Caracas, 1956.

FOGOLARI G., MARCELLO A. — *Su di un'ara altinate.* Estratto: dagli « Atti dell'Istituto Veneto di Scienze, Lettere ed Arti ». Anno accad. 1958-59. Tomo CXVII. Cl. di scienze morali e lettere. Venezia, Officine grafiche. Carlo Ferrari. 1959.

MATEMATICA

AMATO VITTORIO. — *Un metodo pratico per la inversione della matrice di Leontief.* Estratto dal fascicolo, n. 3 de « L'Industria ». Editrice l'Industria, Milano, 1959.

FERRARA MORI GINA. — *Contributo allo studio dell'apprezzamento di lunghezze lineari* in « Atti della Fondazione Giorgio Ronchi ». Anno XIII, n. 1, gennaio-febbraio 1958, Tip. Giuseppe Bruschi, Firenze '58.

FINCH P. D. — *Notes on a combinatorial problem of B. O. Koopman.* N. 85, « Operational Research Quarterly ». Vol. 9, No. 3, 1958.

GUMBEL E. J. — *The M^{th} range* dal « Journal de Mathematiques ». Tome 39, Fascicule N. 3, Gauthier-Villars, Paris, 1959.

SAUVY ALFRED. — *Sur la notion de capital accumulé.* Congrès de l'Association Internationale pour l'étude du revenu et de la Fortune, Pietersberg, 18-24 Août 1957.

STERIOTIS P. J. — *Limite de la valeur d'un capital placé à intérêt composé* Athènes, 1959.

Tables of the bivariate normal distribution function and related functions. — National Bureau of Standards Applied Mathematics Series. 50. Issued June 15, 1959. U. S. Government Printing Office, Washington 25, D.C.

MEDICINA

- ALSTROM C. H., HALLGREN B., NILSSON L. B. and ASANDER H. — *Retinal degeneration combined with obesity, diabetes mellitus and neurogenous deafness* in: «Acta Psychiatrica et Neurologica Scandinavica». Supp. 129, Vol. 34, 1959, Ejnar Munksgaard - Copenhagen, 1959.
- BRUN KETIL. — *Drinking Behaviour in small Groups*. Alcohol Research in the Northern Countries. The Finnish Foundation For Alcohol Studies. Printed in Finland. Helsinki, 1959.
- CROXTON FREDERICK E. — *Elementary Statistics with Applications in Medicine and the Biological Sciences*. Dover Publications, Inc., New York, 1953.
- FORSSMAN O. and LINDEGARD B. — *The post-coronary patient - A multi-disciplinary investigation of middle-aged Swedish males* in: «The Journal of Psychosomatic Research», Vol. 3, Pergamon Press Ltd. London, 1958.
- FRANCESCHETTI-BASEL A. — *Die Vererbung von Augenleiden* in: «Kurzes Handbuch der Ophthalmologie. Erster Band, Anatomie - Entwicklung Missbildungen-Vererbung.
- GAJDUSEK D. C. and ZIGAS V. — *Kuru - Clinical, Pathological and Epidemiological Study of an Acute Progressive Degenerative Disease of the Central Nervous System among Natives of the Eastern Highlands of New Guinea* in: «The American Journal of Medicine». Vol. XXVI, No. 3, 1959.
- GAJDUSEK D. C., ROGERS N. and BANKHEAD A. S. — *Serologic survey of viral and rickettsial diseases among jungle inhabitants of the upper Amazon basin* in «Pediatrics». Vol. 23, No. 1, Part I, January 1959.
- GAJDUSEK D. C. and ZIGAS V. — *Degenerative disease of the central nervous system in New Guinea. The Endemic Occurrence of «Kuru» in the Native Population* in: «New England Journal of Medicine», November 14-1957, Department of Health Education, and Welfare, Public Health Service.
- GAJDESEK D. C. and ZIGAS V. — *Untersuchungen über die pathogenese von Kuru* in: «Klinische Wochenschrift». Heft., 10 mai 1958.
- GOLDSMITH K. L. G. (Editor). — *Blood Groups*. British Medical Bulletin. Published by the Medical Department, The British Council, Vol. 15, No. 2, May 1959.
- GOODMAN HAROLD O. and HERNDON NASH C. — *Genetic Factors in the Etiology of Mental Retardation* in: «International Record of Medicine». Vol. 172, No. 2, Feb. 1959.
- McILWRAITH T. F. — *Frazer of «the Golden Bough»* in: «The Canadian Medical Association Journal», No. 74, 1956, Canada, 1956.
- KLATZO I. and GAJDUSEK D. D. and ZIGAS V. — *Pathology of Kuru*. Paper Presented at the Annual Meeting of the American Association of Neuropathologists, Atlantic City, June 15-16 1958.
- LINNEROTH KERSTIN, HALLGREN BERTIL, PALMEN KURT and ZETTERSTROM ROLF. *Primary (Genuine) Vitamin D Resistant Rickets IV. A Clinical and Genetic Study* in: «Acta Pædiatrica», 47, Sept. 1958.
- MANUILA ALEXANDER. — *Blood groups and disease-hard facts and delusions* in: «The Journal of the American Medical Association». Vol. 167, No. 17 (August 23 1958).

- NEEL JAMES V. — (Editorial) *Medicine's genetic horizons* in : « Annals of Internal Medicine ». Vol. 49, No. 2, August 1958.
- SCHAEFFER M., GAJDUSEK D. C., BROWN LEMA A. and EICHENWALD. — *Epidemic jungle fevers among okinawan colonists in the bolivian rain forest*. I. Epidemiology in : « American of Tropical Medicine and Hygiene ». Vol. 8, No. 3, May 1959.
- Sewage Contamination of Bathing Beaches in England and Wales*. Privy Council - Medical Research Council. Memorandum N. 37, 1959. Her Majesty's Stationery Office London, 1st December 1959.
- SHAPIRO HARRY L. — *Youthful Offenses in Modern Culture*. Reprinted from Bulletin of the New York Academy of Medicine. Second Series, Vol. 35, No. 6, June 1959.
- STENSTEDT AKE. — *Involuntional melancholia* in : Acta Psychiatrica et neurologica Scandinavica ». Suppl. 127, Vol. 34, 1959, Ejnar Munksgaard, Copenhagen 1959.
- ZIGAS V. and GAJDUSEK D. C. — *Kuru : clinical study of a new syndrome resembling paralysis agitans in natives of the eastern higlands of Australian New Guinea*. Department of Health Education, and Welfare, Public Health Service.
- ID. — *Kuru* in : « Papua and New Guinea Medical Journal ». Vol. 3, No. 1, Feb. 1959.

PEDAGOGIA

- FORTUNATI PAOLO. — *Scuola e politica scolastica*. Tip. del Senato « G. Bardi », Roma, 1958.

POLITICA

- BIDWELL PERCY W. — *Raw materials*. A Study of American Policy. Oxford University Press, Oxford, 1958.
- BUCHMAN FRANK. — *Ideologia e Coesistenza*. Citazioni tratte dalla raccolta dei suoi discorsi, « Rifare il Mondo » editore Sansoni, 1959 Riarino Morale Mountain House Caux Svizzera.
- CEVAT ACIKALIN MEHMET. — *La Turquie devant les problèmes internationaux*. A cura del Banco di Roma. Roma, 1959.
- FERRARI AGGRADI MARIO. — *Per una Politica di Sviluppo*. Discorso al VII Congresso Nazion. della Democrazia Cristiana., 1959, ATEL, Roma
- FORTUNATI PAOLO. — *Forma e sostanza del Piano della Scuola*. Discorso pronunciato al Senato della Repubblica nella seduta del 15 ottobre 1959. Tipografia del Senato, Roma, 1959.
- NIXON RICHARD M. — *I Discorsi di Mosca* : 1) 24 luglio 1959 ; 2) 1 agosto 1959. United States Information Service. Stamp. Apollon, Roma 1959.
- SPAAR PAOLO ENRICO. — *Le ragioni della NATO*, Edizioni 5 Lune, Roma, 1959.
- UNITED STATES INFORMATION SERVICE IN ITALY (April 4, 1959). — *Nato and the Atlantic Community 1949-1959*. A Selective List of Publications and Films. Tip. Carpentieri, Roma, 1959.
- ZINGONE VITTORIO. — *L'allarme viene dal Sud* in : « L'Europeo », 19 aprile 1959, Milano, 1959.
- WEBSTER RICHARD A. — *La Rinascita della democrazia cristiana in Italia (1929-1945)*. Estratto da : « Il Mulino ». Bologna, Agosto 1959.

POLITICA ECONOMICA

- CAMAITI ROMOLO. — *Su di uno strumento di largo impiego nel campo economico: gli indici di concentrazione e le loro coincidenze* in: « Rivista di Politica Economica ». Roma, Nov. 1958.
- ERHARD LUDWIG. — *Questioni attuali della politica economica*. A cura del Banco di Roma, Roma, 1959.
- FERRARI AGGRADI MARIO. — *Le Partecipazioni statali nella politica di sviluppo*. Ministero delle Partecipazioni Statali. Roma, MCMLIX.
- MILHAUD MAURICE. — *The Historical development of Social policy*. Uitgeverij van Keulen N. V., 1958. Publications on Social Shange. Number 7.
- MIRABELLA GIUSEPPE. — *Politica monetaria*. Seminario di Economia e Scienza delle Finanze dell'Università degli Studi di Palermo. Palermo, 1959.
- ID. — *Tensioni di divario economico nella comunità economica europea*. Seminario di Economia Politica e Scienza delle Finanze dell'Università degli Studi di Palermo. Palermo, 1959.
- PARRILLO FRANCESCO. — *Politica di sviluppo del credito all'artigianato*. Estratto: Dal Fasc. VII, Luglio 1959, Anno XLIX, III Serie, Rivista di « Politica Economica ».
- RESEARCH TECHNIQUES DIVISION. — *Report on the first ten Years*. No. 103 in the reprint Series of the Division. The Research techniques division of the London School of economics and political Science., 1959.
- VIANELLI SILVIO. — *La Sicilia nel primo ventennio del secolo XX*. Dal « Bollettino dell'Ufficio Studi della Cassa di Risparmio per le Province Siciliane. Palermo, Sett. '59.

POPOLAZIONE

- People's Republic of China*. Guide de Chine. Ed. Langues Etrangères. Pekin, avril 1958.
- MANUILA SABIN, FILDERMAN W. — *Regional development of the Jewish population in Romania*. Estratto da: « Genus » Vol. XIII, No. 1-4, 1957. Tip. F. Failli, Roma.
- ISHWARAN K. — *Family Life in the Netherlands*. Uitgeverij van Keulen, N. V., 1959.
- DUNN L. C. — *Heredity and Evolution in Human Populations*. Cambridge, Massachusetts, 1959, Harvard University Press.
- BOLOMBO BERNARDO. — *Dizionario demografico multilingue*. Volume italiano. Union Internationale pour l'étude scientifique de la population, Editore Giuffrè, Milano, 1959.
- CIASUTTI R. — *Razze e popoli della terra*. 3ª edizione:
Vol. I - *Razze, popoli e culture*.
Vol. II - *Europa e Asia*.
UTET, Torino, 1959.

PSICOLOGIA SOCIALE

- ROSER O. L., NICOLETTI I. — *Contributo allo studio dell'affaticamento mentale nella scuola media* in: « Archivio di psicologia neurologia e psichiatria ». Luglio-agosto '56, Anno XVII, Fasc. IV.
- CHIARI S. — *Rilievi quantitativi e differenziali su protocolli roschach di 108 fanciulli fiorentini dai 6 ai 10 anni* in: « Archivio di psicologia neurologia e psichiatria ». Maggio-agosto 1957, Anno XVIII, Fasc. III-IV.

RICERCA OPERATIVA

- FINCH P. D. — *Balking in the Queueing System GI/M/1*. No. 100 in the Reprint Series of the Research techniques division London school of economics and Political Science. This paper appeared in «Acta Mathematica». Vol. 10, Parts 1 and 2, 1959.
- KENDALL M. G. — *The Teaching of operational research*. No. 99 in «Operational Research Quarterly». Vol. 9, No. 4, Dec. 1958.
- MORTON G. — *Linear programming. An application in a Indian textile mill*. No. 86, in «Operational Research Quarterly». Vol. 9, No. 3, 1958.

SCIENZA DELLE FINANZE

- AUKRUST ODD and BJERKE JUUL. — *Real Capital in Norway 1900-1956*. Fifth Conference - De Pietersberg 19-25 August 1957. International Association for Research in Income and Wealth (The Central Bureau of Statistics of Norway).
- PANCIERA EMILIO. — *Autogoverno amministrativo e autosufficienza finanziaria degli Enti Locali nella nuova legge regionale siciliana* in: «Archivio finanziario», Vol. VI, CEDAM, Padova, 1957.
- ID. — *L'insegnamento della Contabilità di Stato nell'ordinamento superiore* in «Stato Sociale». N. 7, 1958, Unione Tip. Editrice, Torino, 1958.

SCIENZE NATURALI

- GULF ITALIA. — *Petrolio di Ragusa 1958*. A.G.A., Roma, 1959.
- MATTEI ENRICO. — *Considerazioni sui problemi internazionali del petrolio*. A cura del Banco di Roma, Roma - 1959.
- MESTRE ARTIGAS CRISTOBAL D. — *El gas carbonico de las fermentaciones vinicas y sus aplicaciones*. Memorias de la Real academia de ciencias y artes de Barcelona. N° 662, Vol. XXXII, N° 14, Barcelona, 1958.
- MONTI MARIO. — *Il Petrolio in Libia*. Estratto dalla Rassegna «Idrocarburi». N. 8, Agosto 1959, Tipografia P.U. G., Roma.
- P. L. — *Legislazione (La) per le ricerche petrolifere in Libia e quella italiana*. Estratto dal fascicolo di Ottobre 1959 della Rassegna «L'Industria Mineraria». Stab. Grafico Fratelli Lega S. a r. l., Firenze, 1959.
- RENZONI MARCELLA. — *Introduzione alla storia naturale di Buffon*. Il terzo volume della Collana dei classici della scienza. Boringhieri Paolo, Torino, 1959.

SOCIOLOGIA.

- ACOSTA CECILIA. — *Cosas sabidas y cosas por saberse* in: «Revista Nacional de Cultura». Settembre-ottobre 1958, Caracas, 1958.
- AFFRICANO RAOUL. — *Parlo in difesa del buon senso*. Editrice «Adriana», Roma, 1958.
- ALEXANDER-KATZ E. — *Subsidies and Grants* in: Public Subsidies in Israel.

- BARBANO FILIPPO, VITERBI MARIO. — *Bibliografia della Sociologia Italiana (1948-1958)* Con prefazione di Renato Treves. Pubblicazioni dello Istituto di Scienze Politiche dell'Università di Torino. Edizioni Ramella, 1959. Torino.
- BOARD OF MEDIATION. — *State of New York Department of Labor. A Guide to. Mediation and Arbitration.* New York State.
- BARENTS Mr. Dr. J. — *Democracy: an unagonized reappraisal.* Uitgeverij van keulen N. V., 1958. Publications on Social Shange. Number 10.
- BEALE E.M.L. and MORTON G. and LAND A. H. — *Solution of a purchase-storage programme.* No. 84 in: «Operational Research Quarterly» Volume 9, No. 3, 1958.
- BELSON W. A. — *Measuring the effects of televisions a description of method* No. 81 in: «Public Opinion Quarterly», Vol. 22, No. 1, 1958.
- ID. — *Research for programme planning in Television* No. 83 in: «ATV Series of Technical Research Studies, 1958.
- BEN-ZVI ITZHAK. — *The exiled and the redeemed.* The Jewish Publ. Society of America Philadelphia, 1957.
- BEN-ZVI INSTITUTE. — *Studies and reports.* II, Jerusalem, 1956.
- CAVALLI LUCIANO. — *La gioventù del quartiere operaio.* Le inchieste dello ufficio studi sociali e del lavoro del comune di Genova, N. 3, Genova, 1959.
- COMAS JUAN. — *Critica científica y espiritu universitario.* Mexico, 1959.
- CORSO RAFFAELE. — *Il «Contrasto della bianca e della bruna» dal punto di vista sociologico* in: «Revue Internationale de Sociologie», Vol. I, N. 2-3. Società Italiana di Sociologia, Roma.
- D'ADDIO MARIO. — *Brevi considerazioni sulla sociologia in Italia.* «Sociologia». Milano, Gen.-Mar. 1959.
- DE SMET E. et EVALENKO RENÉ. — *Les élections Belges: Explications de la répartition géographique des suffrages.* Collection de science politique. Université Libre de Bruxelles Institut de Sociologie Solvay, Bruxelles, 1956.
- DI CARLO E. — *La sociologia in Italia dalla seconda metà dello scorso secolo ad oggi.* «Sociologia». Milano, Gen.-mar. 1959.
- ID. — *La sociologia in Italia dalla metà dello scorso secolo ad oggi* in «Sociologia». Lug.-Set., 1958.
- DUPUY PIERRE. — *Dalla Torre d'Avorio allo Sputnik* (Discorso pronunciato a Roma, il 5 marzo 1959, nella Sede del Banco di Roma. (A cura del Banco di Roma). Stab. A. Staderini, Roma.
- EVERSLEY D. E. C. — *Social theories of Fertility and the Malthusian debate.* Oxford at the Clarendon Press. 1959.
- GRONCHI GIOVANNI. — *Per l'unità e la solidarietà tra i popoli.* A cura del Banco di Roma. Roma, 25 nov. 1958.
- MANCINELLI GIUSEPPE. — *L'assenza del problema militare odierno.* A cura del Banco di Roma. Roma, 29 aprile 1959.
- MAROTTA MICHELE. — *Organicismo e Neo-organicismo.* Estratto da «Sociologia» bollettino dell'Istituto Luigi Sturzo. Anno IV, Aprile-Giugno 1959, N. 2, Casa Editrice A. Giuffrè, Milano, 1959.
- ID. — *Organicismo e Neo-organicismo.* Estratto da «Sociologia», bollettino dell'Istituto Luigi Sturzo. Anno IV, Gennaio-marzo 1959, N. 1, Casa Editrice A. Giuffrè, Milano, 1959.

- ID. — *Migratorie, correnti, mafia* in : « Grande dizionario Enciclopedico ». Vol. VIII, Unione Tip. Editrice torinese, Torino, 1958.
- ID. — *Il pensiero sociologico di Corrado Gini e la sociologia in Italia : A proposito di una recente ristampa del suo corso di sociologia*. Note e discussioni in : « Sociologia ». Lug.-Set., 1958.
- MENDIETA Y MUÑEZ LUCIO. — *Teoría de la revolución*. Cuadernos de Sociología. Instituto de Investigaciones Sociales, Universidad Nacional, Mexico, 1959.
- MULLER KARL VALENTIN. — *Soziologische und sozialpsychologische Folgen der Austreibungen* in : « Das Ostliche Deutschland ».
- ID. — *Zür Wandlung der sozialen prestigeordnung unter sozialistischem einfluss* in : « Zeitschrift für die gesamte statswissenschaft 115 Band/ 2 Heft J. C. B. Mohr (Paul Siebeck) Tubingen, 1959.
- NIRCHIO GIUSEPPINA. — *A proposito di vecchia e nuova sociologia* in : « Il politico ». Anno XXIII, N. 4, 1958, Casa ed. Giuffrè, Milano, 1959.
- NIRCHIO GIUSEPPINA. — *A proposito di vecchia e nuova sociologia* in « Il Politico », Dic. 1958. Presso l'Università degli studi di Pavia, Dic. 1958.
- OSBURN WILLIAM F. and NIMKOFF MEYER F. — *Sociology*. Third ed. Houghton Mifflin Company. Boston 1958.
- PADOAN GIOVANNI. — *Sullo spopolamento montano. Realtà del fenomeno e possibili rimedi* in : « Lavoro e Sicurezza Sociale ». Roma, maggio 1959.
- PEDRONI F. — *Aspetti umani e sociali della produttività per il miglioramento del tenore di vita*. — Articolo tratto da : il giornale « Azione Sindacale », Pagina, 3. Stab. Tipog. U.E.S.I.S.A., Roma.
- ROSE ARNOLD M. — *Indagine sull'integrazione sociale in due quartieri di Roma*. Istituto di Statistica Sezione di Sociologia - Centro di Ricerche di Sociologia Empirica. Nuova Serie : Sezione di Sociologia Studi e inchieste n. 1, Stab. Tipog. Failli, Roma, 1959.
- RUNES DAGOBERT D. — *On the nature of man*. Philosophical Library, New York, 1956.
- SPROTT W. J. H. — *Sociology*. Hutchinson University Library. London, 1959.
- TRUJILLO CARLOS A. ECHANOVE. — *La Radiodifusion y la cultura*. Estudios de Sociologia Empirica. IV Consejo Superior de Investigaciones Cientificas. Instituto « Balmes » de Sociologia Madrid, 1959.
- UNITED NATIONS EDUCATIONAL SCIENTIFIC AND CULTURAL ORGANIZATION. *Census data on level of education*. Unesco, Paris 1958.
- UNITED STATES INFORMATION SERVICE. — *Berlino*. Uno studio analitico. Roma, gennaio 1959.
- THE UNIVERSITY OF CEYLON. — *The disintegrating Village*. A socio-economic survey conducted. Part. I. The Ceylon University Press Board 1957.
- VOGT EDVARD. — *Introduzione alla sociologia*. Università internazionale degli Studi - Istituto di sociologia. Roma, Anno Accademico, 1958-59. Ed. Internazionale sociali. Roma, 1959.
- ID. — *Le leggi sociologiche*. A. Giuffrè Editore, Milano 1958.

SOCIOLOGIA : ABITAZIONI

- BOGUE DONALD J. — *A system of socio-economic regional boundaries for the americas.*, ISI, Stockholm, 1957.

SOCIOLOGIA BIOPSIKOLOGICA

BOSCHI GAETANO. — *Sociologia Biopsicologica*. Accademia di Scienze Lettere e Arti di Modena. Società Tip. Editrice modenese. Modena, 1957.

STATISTICA AGRICOLA

NATIONS UNIES. — *Situation de l'électrification rurale en Europe en 1957*. Genève, 1959.

GAETANI ALFONSO. — *L'Italia agricola e il mercato comune europeo*. A cura del Banco di Roma. Roma, 1958.

STATISTICA APPLICATA

BANDETTINI PIERFRANCESCO. — *I Numeri indici*. Serie di statistica - Teoria e Applicazione. Edizioni Scientifiche, Einaudi, 1959.

CRAMER J. S. — *The depreciation and mortality of motor cars* N. 145 in: « The Journal of the Royal Statistical Society, Series A., Vol. 121, Part. I, 1958.

GALES KATHEEN. — *A survey of fellows* - I No. 90 in: « Journal of the Royal Statistical Society », Serie A, Vol. 121, Part. 4, 1958.

GRADARA ENRICO. — *La rappresentazione grafica dei fenomeni statistici*. Serie di statistica - Teoria e applicazioni, 10. Edizioni Scientifiche Einaudi, Torino, 1959.

KRUSKAL WILLIAM H. — *Recensione di: Nonparametric and Shortcut Statistics*. Merle W. Tate and Richard C. Clelland. Danville Illinois: Interstate Printers and Publishers, Inc. 1957 in: « Journal of the American Statistical Association, June, 1958, Vol. 53.

ROSENBERG LOUIS. — *Jewish children in the Protestant schools of Greater Montreal in the period from 1878 to 1958. A statistical study* in: « Research Papers ». S. E., No. 1, March 16-1959, Bureau of Social and economic Research Canadian Jewish Congress

WOLD HERMAN. — *Causal Inference from Observational Data a Review of Ends and Means* in: « The Journal of the Royal Statistical Society ». Serie A, Vol. 119, Part. 1, 1956. Printed for Private Circulation 1956.

STATISTICA AZIENDALE

ISTITUTO DI STATISTICA (Università di Bologna - Centro Meccanografico). — *Relazione sulle entrate e sulle spese degli istituti, delle scuole di perfezionamento, dei centri studi e delle gestioni speciali nell'anno accademico 1955-56*. Tip. Compositori, Bologna.

PANCIERA EMILIO. — *Introduzione allo Studio dell'Economia Aziendale*. Tip. Luxograph - Palermo, 1955.

ID. — *A proposito di alcuni corsi di addestramento per dirigenti di azienda* in: « Annali della facoltà di economia e commercio ». Anno XII, 1958, N. 1, Abbaco Editore, Palermo, 1958.

ID. — *Le macroaziende*. Vol. I - L'Azienda Statale. Abbaco Editore, Palermo, 1957.

STATISTICA ECONOMICA

- ALBERTI SALVATORE. — *Criteri per la costruzione di un indice generale dei prezzi dei beni e dei servizi* in: « Rivista Internazionale di Scienze Economiche e Commerciali » Anno IV (1957), No. 10, CEDAM, Padova, 1957.
- BANDETTINI PIERFRANCESCO. — *Statistica economica*. Le fonti per lo studio dell'economia Italiana. 2^a ed., CEDAM, Padova, 1959.
- GRADARA ENRICO. — *La Rappresentazione grafica dei fenomeni statistici*. A cura del servizio Statistica e Studi economici e del Servizio Centrale meccanografico FIAT. Edizioni Scientifiche Einaudi, Torino, 1959.
- HIRSCH WERNER Z. — *Introduction to Modern Statistics with applications to Business and Economics*. The MacMillan Co., New York, 1957.

STATISTICA METODOLOGICA

- BACHI ROBERTO. — *Statistical analysis of geographical series*. Kaplan School, Hebrew University and Central Bureau of Statistics, Jerusalem.
- BARBERI BENEDETTO. — *Rilevazioni statistiche*. Serie di Statistica - Teoria e applicazioni. Edizioni scientifiche Einaudi. Torino, 1957.
- ID. — *Some applications of the sampling method in Italian official statistics*. The International Statistical Institute, Stockholm, 8-15 August 1957. Central Institute of Statistics, Roma, 1957.
- BERGE CLAUDE. — *Théorie des graphes et ses applications*. Dunod, Paris, 1958.
- BIGNARDI FRANCESCO. — *Introduzione alle applicazioni statistiche di particolari criteri di analisi infinitesimale*. Parte Prima. Estratto da: « Annali della Facoltà di Economia e Commercio dell'Univ. di Palermo ». Anno XIII, 1959, N. 2, Abbaco Editore, Roma.
- CHAPMAN DOUGLAS G. — *A comparative study of several goodness-of-fit tests* in: « The Annals of Mathematical Statistics, Vol. 29, No. 3, Sept. 1958.
- COPPINI MARIO ALBERTO. — *Sull'Applicazione dell'Analisi delle Medie alla Tecnica Attuariale*. Estratto: dalla « Rivista Italiana di Economia, Demografia e Statistica ». Vol. XII, 1958, Soc. Abete, Roma.
- DE WOLF P. — *On the application of stratified sampling to an auditing problem*. (Estratto dal volume in onore di Corrado Gini). Tip. Fausto Failli, Roma.
- DANIELS H. E. and KENDALL M. G. — *Short proof of miss Harley's theorem on the correlation coefficient* No. 92,93,94. In « Biometrika », Vol. 45, Parts. 3-4, Dec. 1958.
- FISHER G. R. — *Maximum likelihood estimators with heteroscedastic errors* No. 146 in: « Review of the International Statistical Institute », Vol. 25, No. 1/3, 1957.
- FORTUNATI PAOLO. — *Statistica e ricerca scientifica* in: « Statistica ». Lug.-Set. 1958, Bologna, 1958.
- FRANZBLAU ABRAHAM N. — *A Primer of statistics for Non Statisticians*. Harcourt, Brace and Company. New York, 1958.
- EIDENOSSISCHES STATISTISCHES AMT. — *Turnprüfung bei der rekrutierung* 1957. Beiträge zur Schweizerischen Statistik - heft 30, Bern 1959.
- VAN ELTEREN PH. — *On the combination of independent two sample tests of Wilcoxon* No. 28. Stichting Mathematisch Centrum. Amsterdam, 1958.

- GOODMAN LEO A., KRUSKAL W. H. — *Measures of Association for Cross Classification. II: Further discussion and References*. Reprinted from «The Journal of the American Statistical Association». March 1959, Vol. 54, pp. 123-163.
- GOODMAN LEO A. — *Asymptotic Distributions of «Pisquared» Goodness of Fit Criteria for m -Th order Markov Chains*. Reprinted from «The Annals of Mathematical Statistics». Vol. 29, No. 4, December, 1958, Printed in U.S.A.
- ID. — *On some Statistical Tests for MTH order Markov Chains*. Reprinted from «The Annals of Mathematical Statistics». Vol. 30, No. 1, March 1959, Printed in U.S.A.
- ID. — *Simple Statistical Methods for Scalogram Analysis* in: «Psychometrika». Vol. 24, No. 1, March 1959.
- GUMBEL E. J. — *Théorie Statistique des débits d'étiage. The Statistical Theory of low water discharges*. Extrait de «La Houille blanche». Numero 1, Janvier-fevrier 1959, Grenoble (Isère).
- HYRENIUS HANNES, HUGOSSON KURT, GUSTAFSSON SVEN-EVERT. *Näringsliv och befolkningsutveckling inom östersunds-regionen* in: «Statistiska Institutionen, Göteborgs Universitet Skriftserie. Publications, Göteborg, 1959.
- JAFFE A. J. — *Handbook of Statistical Methods for Demographers*. Selected Problems in the Analysis of Census Data. (Preliminary Edition - Second Printing). United States Department of Commerce, 1951.
- KENDALL M. G. — *Research in Statistics*. No. 80.
- KENDALL M. G. — *Sampling as an exact science*. No. 88, in «Lloyds Bank Review», Jan. 1959.
- KRUSKAL WILLIAM H. — *Ordinal Measures of Association*. — Reprinted from the: «Journal of the American Statistical Association». December 1958, Vol. 53, pp. 814-861.
- LOIZELIER BLANCO. — *La inspeccion estadistica en las industrias de tipo medio*. L'Institut International de Statistique. Stockholm, 1957.
- MCCARTHY PHILIP J. — *Introduction to Statistical Reasoning*. McGraw - Hill House, London, EC4.
- MENDENHALL W. and HADER R. J. — *Estimation of parameters of mixed exponentially distributed failure time distributions from censored life test data*. No. 96 in «Biometrika». Vol. 45, Parts. 3-4, Dec. 1958.
- MENDENHALL W. — *A bibliography on life testing and related topics*. No. 95 «Biometrika», Vol. 45, Part. 3-4, Dec. 1958.
- MIANI-CALABRESE DONATO. — *Metodologia statistica e statistica di fenomeni sociali*. Dott. A. Giuffré Editore, Milano, 1958.
- MOORE P. G. — *Principles of Statistical Techniques*. A first Course from the beginnings for Schools and Universities. Cambridge University Press. 1958.
- OLKIN INGRAM and PRATT JOHN W. — *Unbiased estimation of certain correlation coefficients* in: «The Annals of Mathematical Statistics». Vol. 20 No. 1, March 1958.
- ID. — *A multivariate Tchebycheff inequality* in: «The Annals of Mathematical Statistics». Vol. 29, No. 1, march 1958.
- PANIZZON GAETANO — *Il coefficiente di correlazione lineare moltiplica generalizzato*. In «Giornale degli Economisti e Annali di economia». Gennaio-febbraio, 1959 CEDAM, Padova, 1959.

- POMPIJ G. e DALL'AGLIO G. — *Piano degli esperimenti*. Serie di Statistica - Teoria e applicazioni. Edizioni Scientifiche Einaudi, Torino, 1959.
- QUENOUILLE M. H. — *The comparison of correlations in time-series* No. 82 in: « Journal of the Royal Statistical Society ». Series B, Vol. XX, 1958.
- QUENOUILLE M. H. — *Discrete autoregressive schemes with varying time-intervals*. N° 75 in « Metrika », Vol. I, Part. I, 1958.
- QUENOUILLE M. H. — *Tables of Random observations from standard distributions*. This paper appeared in « Biometrika », Vol 46, Parts 1 and 2, June 1959.
- RAMASUBBAN T. A. — *The mean difference and the mean deviation of some discontinuous distributions*. No. 92-93-94. In « Biometrika ». Vol. 45, Parts. 3-4, Dec. 1958.
- SALVEMINI TOMMASO. — *Regressione e Correlazione*. Serie di Statistica. Teoria e Applicazioni. Edizioni Scientifiche Einaudi, 1959.
- STUART A. — *Iterative solutions of likelihood equations*. No. 92-93 and 94. In « Biometric Society ». Vol. 14, No. 1, March 1958.
- STUART ALAN. — *Equally correlated variates and the multinormal integral* No. 89, in: « Journal of the Royal Statistical Society ». Serie B, Vol. 20, Part. 2, 1958.
- TATE R. F. and GOEN R. L. — *Minimum variance unbiased estimation for the truncated Poisson distribution* in: « The Annals of Mathematical Statistics ». Vol. 29, No. 3, Sept. 1958.
- WOLD HERMAN O. A. — *Ends and Means in Econometric Model Building* Reprint from « Probability & Statistics ». The Harald Cramér Volume. Ed. Ulf Grenander, 1959.

STATISTICA SANITARIA

- L'ELTORE GIOVANNI. — *Sull'uso dei rapporti statistici e del rapporto di concentrazione nella statistica sanitaria*. L'INADEL, Aprile 1959, Roma.

STATISTICA: VARIE

- FRIIS HENNING. — *The Danish National Institute of Social Research*. September 1959.

STORIA

- GEISENDORF PAUL-F. — *L'Université de Genève 1559-1959*. A. Jullien, Genève 1959.
- GIGLIO CARLO. *L'Italia in Africa* Serie Storica - Vol. I *Etiopia - Mar Rosso* Tomo I (1957-1885), Ministero degli Affari Esteri, Istituto Poligrafico dello Stato, Roma, 1958.
- NICOLAI GEORG F. — *La Biologia de la Guerra*. Traduccion del Aleman por D. A. De Santillan. Apendice: La guerra y la Soberania de los Estados Mexico, Lima, Buenos Aires.
- SCHOTT FRANCIS H. — *The evolution of latin american exchanhe-rate policies sice world war II*. Essays in International Finance No. 32, January 1959, Princeton University, Princeton, 1959.

STANLEY GEORGE F. G. — *The Indian Background of Canadian History*. The Canadian Historical Association, 1952.

VAN VEEN J. — *Dredge Drain Reclaim. The Art of a Nation*. Martinus Nuhoff, The Hague, 1952.

ZOOLOGIA

SUI-TONG CHAN FUNG — GOWEN JOHN W. — *Pigment-Inducing Potentialities of testes, Ovaries and Hermaphroditic (HR) Gonads*. Estratto: « The Journal of Experimental Zoology », Vol. 135, No. 1, June 1957.

MISCELLANEA

CENTRO PER LO SVILUPPO DEI TRASPORTI AEREI. *VIII Assemblea Generale dei soci* Roma, 19 dic. 1958.

ID. — *L'aviazione civile all'assemblea costituente e nel parlamento del dopo guerra*. Roma, 1959.

ID. — *Problèmes juridiques de l'espace supra-atmosphérique*. Roma, 29 Feb. 1959.

ID. — *Lo « Status » giuridico dell'aeromobile secondo i progetti dell'I.C.A.O.* Conferenza del Prof. Riccardo Monaco. Roma, 6 marzo 1959.

ID. — *Politica dei trasporti e coordinamento del mezzo aereo con quelli terrestri*. Prolusione del Sen. Armando Angelini. Roma, 2 maggio 1959.

ID. — *Trasporti marittimi e mercato comune europeo*. Conf. del Dott. Vito Dante Flore. Roma, 15 maggio 1959.

ID. — *Cooperation between airlines: economic and political aspects*. Conf. del Dr. Lambertus H. Slotemaker. Roma, 22 maggio 1959.

ID. — *Prospects of future air traffic development within the european common market* Conf. dell'Ing. Bo Björkman. Roma, 29 maggio 1959.

ID. — *Prospettive della politica dei trasporti terrestri nel mercato comune europeo*. Conferenza del Prof. Francesco Santoro. Roma, 8 maggio 1959.

CONGRESSO VII ITALIANO DEI TRASPORTI. — *Il sistema dei trasporti in relazione alla evoluzione delle fonti di energia. Programma*. Milano 6-7-8 Dicembre 1959. Museo Nazionale della Scienza e della Tecnica « Leonardo da Vinci ».

FRISCH (VON) DR. KARL. — *Aus Dem Leben der Bienen*. Copyright 1927, 1948, 1953 by Springer-Verlag OHG. Berlin, Göttingen, Heidelberg, Printed in Germany.

NATIONS UNIES. — *Perspectives de l'énergie nucléaire comme Nouvelle source d'énergie*. Genève, 1959.

NATIONS UNIES - CONSEIL ECONOMIQUE ET SOCIAL. — *Rapport sur les problèmes économiques du stockage souterrain du gaz*.

NATIONS UNIES. — *Travaux de la réunion spéciale sur la normalisation et la coordination modulaire dans le bâtiment*. Publications de Nations Unies. Genève, 1959.

NATIONS UNIES - CONSEIL ECONOMIQUE ET SOCIAL. — *Utilisation de l'Espace dans les logements*. Août 1959, Nations Unies, Genève, 1959.

NATIONS UNIES. — *Le Logement en Europe en 1958 Tendances et Politiques*. Nations Unies, Genève, 1959.

MISCELLANEA : COMMITTEE OF PRIVY COUNCIL FOR MEDICAL RESEARCH

DREW G. C., COLQUHOUN W. P., PH. D. and HAZEL A. LONG. — *Effect of Small Doses of Alcohol on a Skill Resembling Driving*. Privy Council. Medical Research Council Memorandum. No. 38. London. Her Majesty's Stationery Office, 1959.

Report of the Medical research council. For the year 1957-58. Presented to Parliament by the Lord President of the Council by Command of Her Majesty. June, 1959. London, Her Majesty's Stationery Office.

MISCELLANEA : INTERNATIONAL CONGRESS OF GENETICS - X.

FREIRE-MAIA N., FREIRE-MAIA A. — *Effects of the Structure of First Cousin Marriages on the Coefficient of Inbreeding for Sex-linked Genes in Brazil*. Reprinted from X International Congress of Genetics, Proceedings, Vol. II. Printed in Canada.

FREIRE-MAIA A., FREIRE-MAIA N., QUELCE, SALGADO A. — *Genetic Aspects of Achiropody*. Reprinted from X International Congress of Genetics, Proceedings, Vol. II, Printed in Canada.

GOWEN JOHN W., MAGNHILD UMAERUS. — *Biological Recovery from Radiation effects as related to Genetics*. Reprinted from X International Congress of Genetics, Proceedings Vol. II, Printed in Canada.

STADLER JANICE, GOWEN JOHN W. — *Implication of Genetics to Radiation Effects on Quantitative Characters*. Reprinted from X International Congress of Genetics, Proceeding Vol. II, Printed in Canada.

MISCELLANEA : 31^{ème} SESSION DE L'INSTITUT INTERNATIONAL DE STATISTIQUE

BRICHLER M. — *Comparabilité de certaines caractéristiques socio-économiques dans les recensements de 1960 en Europe* No. 39. 31st Session of the International Statistical Institute in Brussels 2-8 sept. 1958.

DIRECTORATE OF STATISTICS MINISTRY OF NATIONAL PLANNING DAMASCUS. — *A brief sketch of the history and present status of the directorate of statistics*. 31st Session of the International Statistical Institute, 25 August 1958.

KOLLER S. — *The Use of Prior Statistical Informations in Problems of Estimation*. Beitrag zur 31. Tagung des International Statistical Institute. Brussels, 3-8 Sept. 1958.

OLSSON WILLIAM W. — *A project for mapping the population of the world-in the scale of 1 : 1 000 000*. 31st Session of the International Statistical Institute in Brussels Sept. 1958.

WIGHTMAN DE CARVALHO ALCEU VICENTE. — *La population du Brésil dans le passé, le présent et l'avenir*. Institut Brésilien de Géographie et statistique. 31^{ème} Session de L'Institut International de Statistique. Bruxelles sept. 1958.

MISCELLANEA : CONGRESSI DI SOCIOLOGIA

BUSTINZA UGARTE PEDRO et SOPEÑA IBAÑEZ ANGEL. — *Análisis de la natalidad española y de su decrecimiento*. Communication au XVII Congrès de Sociologie.

IBÁÑEZ ESTEBAN. — *Personalidad de la mujer en la sociedad bereber*. Communication au XVII Congrès de Sociologie.

YAGUE JUAN GARCIA. — *La ruptura familiar como problema para la infancia*. Institut International de Sociologie, XVIII Congrès, Nuremberg 10-17 sept. 1958.

MISCELLANEA : 6th CONGRESS OF THE EUROPEAN SOCIETY OF HAEMATOLOGY.

MOOR-JANKOWSKI JAN K., HUSER H. J. and GEIGER M. — *Haemophilia B in the Wide-Spread Kindred of tenna Haemophiliacs* in : « Transactions of the 6th Congress of the European Society of Haematology » Copenhagen, 1957.

MISCELLANEA DI SOCIOLOGIA

Atti del Seminario di Sociologia. Anno Accademico 1955-56. Istituto di Statistica Sezione di Sociologia - Centro Ricerche di Sociologia Empirica. Vol. I - Nuova Serie. Stab. Tipog. Failli, Roma.

MISCELLANEA : COMPTES-RENDUS DES SEANCES DE L'ACADEMIE DES SCIENCES

LEJEUNE M. JÉRÔME, GAUTIER M. M., TRPIN RAYMOND. — *Les chromosomes humains en culture de tissus*. Comptes-rendus des séances de l'Académie des Sciences, T. 248, 26 janvier 1959.

Id. — *Etude des chromosomes somatiques de neuf enfants mongoliens* in : « Comptes-rendus des séances de l'Académie des Sciences. T. 248, Séance 16, mars 1959.

MISCELLANEA : 5^a e 6th. CONFERENCE INTERNATIONAL ASSOCIATION FOR RESEARCH IN INCOME AND WEALTH

BICANIC RUDOLF. — *Economic growth by Economic sectors*. International Association for Research in income and wealth. 6th Conference, Portoroz., August 23, September 1st 1959.

BJERVE PETTER JACOB and SELSJORD MIKAEL. — *Financial accounting within a system of National Accounts* The Central Bureau of Statistics of Norway. Fifth Conference De Pietersberg 19-25 Agosto 1957. International Association for research in income and wealth.

MISCELLANEA : ITALIAN COMMITTEE FOR HEALTH EDUCATION

THE ITALIAN COMMITTEE FOR HEALTH EDUCATION. — *Third International Conference on health education. Rome 28 april, 5 may 1956.* Official Proceedings, Vol. I-II, Roma, 1956.

MISCELLANEA : INST. INTER. DE STATIST.

BONOMI PAOLO :

Vol. I. *Parte generale.*Vol. II. *Documentazione statistica.*

in : « XIII Congresso Nazionale, 22-23-24 aprile. Confederazione nazionale dei coltivatori diretti. Roma, 1959.

MISCELLANEA : 2^e CONGRES DES ECRIVAINS ET ARTISTES NOIRSL'UNITÉ DES CULTURES NEGRO-AFRICAINES. — *Présence africaines*. Deuxième Congrès des Ecrivains et Artistes Noirs. Rome, 26 mars-1^{er} avril 1959.

VARIE

Bastions de Genève. Numéro Special du IV^e Centenaire. No. 4 - Deuxième semestre, 1959. Revue semestrielle publiée par l'Association des anciens étudiants de l'Université de Genève. 1959.FIDEL E. RAURICH D. — *Labor de un academico profesor de la facultad de farmacia*. Memorias de la Real academia de ciencias y artes de Barcelona, N. 663, Vol. XXXII, N. 15. Barcelona, 1958.HECKER HERMANN. — *Zür Geschichte der Landesplanung Über sozialen Wohnungsbau, Städtebau und Bauberatung zur Landesplanung*. Hammonia-Verlag GmbH Hamburg 1959.KUNSTLERGILDE E. V., SITZ ESSLINGEN N. — *Stimmen aus dem schatten*. Darmstadt 7 bis, 10 Marz. 1958.NAVILLE PIERRE. — *Proposte di riforme* in : « Longevità », Set.-dic. 1958.PEDRONI FERNANDO. — *Prevedibili conseguenze tecniche e sociali della diffusione delle calcolatrici elettroniche*. — Estratto dalla : « Rivista Italiana di Economia, Demografia, e Statistica ». Vol. XIII, N. 3-4, 1959, Soc. A.B.E.T.E. - Roma.SHAPIRO HARRY L. — « *Exhibit in Progress*. A Quarterly Publication of « Curator » the American Museum of Natural History. Vol. II, No. 3, 1959.*Weight Gains, Serum Protein Levels, and Health of Breast Fed and Artificially Fed Infants*. Special Report Series, N. 296, London : Her Majesty's Stationery Office. 1959.

PUBBLICAZIONI UFFICIALI

Annuaire du collège de France. 59^e Année. Imprimerie nationale, Paris, 1959.*Anuarul statistic al R.P.R.*, 1958. Republica Populara Romina. Directia Centrala de Statistica. Bucaresti, 1958.*Anuarul statistic al R.P.R.*, 1959. Republica Populara Romina. Directia Centrala De Statistica. Bucaresti, 1959.BATALLER CALATAYUD JOSÉ R. — *El cretacico del macizo del Montmell (Terragona)*. Memorias de la Real Academia de Ciencias y artes de Barcelona, N^o 664, Vol. XXXII, N^o 16, Barcelona, 1958.

CULTURAL PROGRESS IN THE U.S.S.R. — Central Statistical Board of the U.S.S.R. - Council of Ministers. Statistical Returns., Printed in the Union of Soviet Socialist Republics., 1958.

DELEGATION DER LIGA DER ARABISCHEN STAATEN (Bonn). — *Veröffentlichungen der Liga der Arabischen Staaten - Kairo*

- Nº 1 Januar 1958
- Nº 2 Februar »
- Nº 3 März »
- Nº 4 April »
- Nº 5 Mai »

DUASON J. — *Landkönnun of Landnam Islendinga I Vesturheimi*, Reykjavik, Prentad I Isafoldarprentsmidju H. F., 1941.

ID. — *Rjettarstada Graenlands Nylendu Islands* ». Reykjavik. Prentad I Isafoldarprentsmidju ILF, 1947

State statistical Bureau of the people's Republic of China - Sept. 21, 1955. Report on National Economic Development and Fulfilment of the State Plan in 1954. Foreign Languages Press, Peking, 1956.

ID. — June 14, 1956. Report on Fulfilment of the National Economic Plan in 1955. Foreign Languages Press, Peking, 1956.

REPUBLICA POPULARA ROMINA. — *Recensamintul Populatiei*, Din 21 Februarie 1956, Resultate Generale. Directia Centrala de Statistica.

ACADEMIE (L) INTERNATIONALE DE PHILOSOPHIE DES SCIENCES.

Symposium 1959.